AN

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 July 2001 (05.07.2001)

PCT

(10) International Publication Number
WO 01/47957 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/US00/35581

(22) International Filing Date:
28 December 2000 (28.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/474,279 29 December 1999 (29.12.1999) US

(71) Applicant: ZYMOGENETICS, INC. [US/US]; 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

(72) Inventor: CONKLIN, Darrell, C.; 117 East Louisa Street #421, Seattle, WA 98102 (US).

(74) Agent: SAWISLAK, Deborah, A.; ZymoGenetics, Inc., 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/47957 A2

(54) Title: NOVEL FGF HOMOLOG zFGF10

(57) Abstract: The present invention relates to polynucleotide and polypeptide molecules for zFGF10, a novel member of the FGF family. The present invention also includes antibodies to the zFGF10 polypeptides, and methods of using the polynucleotides and polypeptides to stimulating growth of cells from the mesenchymal lineage.

Description

NOVEL FGF HOMOLOG zFGF10

BACKGROUND OF THE INVENTION

The fibroblast growth factor (FGF) family consists of at least eighteen distinct members (Basilico et al., Adv. Cancer Res. 59:115-165, 1992 and Fernig et al., Prog. Growth Factor Res. 5(4):353-377, 1994) which generally act as mitogens for a broad spectrum of cell types. For example, basic FGF (also known as FGF-2) is mitogenic *in vitro* for endothelial cells, vascular smooth muscle cells, fibroblasts, and generally for cells of mesoderm or neuroectoderm origin, including cardiac and skeletal myocytes (Gospodarowicz et al., J. Cell. Biol. 70:395-405, 1976; Gospodarowicz et al., J. Cell. Biol. 89:568-578, 1981 and Kardami, J. Mol. Cell. Biochem. 92:124-134, 1990). *In vivo*, bFGF has been shown to play a role in avian cardiac development (Sugi et al., Dev. Biol. 168:567-574, 1995 and Mima et al., Proc. Nat'l. Acad. Sci. 92:467-471, 1995), and to induce coronary collateral development in dogs (Lazarous et al., Circulation 94:1074-1082; 1996). In addition, non-mitogenic activities have been demonstrated for various members of the FGF family. Non-proliferative activities associated with acidic and/or basic FGF include: increased endothelial release of tissue plasminogen activator, stimulation of extracellular matrix synthesis, chemotaxis for endothelial cells, induced expression of fetal contractile genes in cardiomyocytes (Parker et al., J. Clin. Invest. 85:507-514, 1990), and enhanced pituitary hormonal responsiveness (Baird et al., J. Cellular Physiol. 5:101-106, 1987.)

Several members of the FGF family do not have a signal sequence (aFGF, bFGF and possibly FGF-9), and thus would not be expected to be secreted in a classical fashion. In addition, several of the FGF family members have the ability to migrate to the cell nucleus (Friesel et al., FASEB 9:919-925, 1995). All the members of the FGF family bind heparin based on structural similarities. Structural homology crosses species, suggesting a conservation of their structure/function relationship (Ornitz et al., J. Biol. Chem. 271(25):15292-15297, 1996.)

There are four known extracellular FGF receptors (FGFRs), and they are all tyrosine kinases. In general, the FGF family members bind to all of the known FGFRs, however, specific FGFs bind to specific receptors with higher degrees of affinity. Another means for specificity within the FGF family is the spatial and temporal expression of the ligands and their receptors during embryogenesis. Evidence,

suggests that the FGFs most likely act only in autocrine and/or paracrine manner, due to their heparin binding affinity, which limits their diffusion from the site of release (Flaumenhaft et al., J. Cell. Biol. 111(4):1651-1659, 1990.) Basic FGF lacks a signal sequence, and is therefore restricted to paracrine or autocrine modes of action. It has been postulated that basic FGF is stored intracellularly and released upon tissue damage. Basic FGF has been shown to have two receptor binding regions that are distinct from the heparin binding site (Abraham et al.,. EMBO J. 5(10):2523-2528, 1986.)

Within the FGF family, the keratinocyte growth factor proteins (KGF) form a distinct group due to the specificity of the molecules for their target cell. There have been two KGF proteins identified KGF, which is also designated FGF-7; and KGF-2, alternatively designated as FGF-10. KGF proteins have specificity for epithelial cells, unlike other members of the FGF family which have been demonstrated to act as mesenchymal cell growth factors (Rubin et al., Cell Biol. Internatl. 19:399-411, 1995 and Danilenko, Toxicol. Pathol. 27:64-71, 1999).

Members of the FGF family have been shown to play important roles developmentally and in adult tissue. The activities of the family members appear to be promiscuous in some tissues and have tissue-specificity in other cases. The present invention provides a novel member of the FGF family and the uses for these polynucleotides and polypeptides should be apparent to those skilled in the art from the teachings herein.

DETAILED DESCRIPTION OF THE INVENTION

Prior to setting forth the invention in detail, it may be helpful to the understanding thereof to define the following terms:

The term "affinity tag" is used herein to denote a polypeptide segment that can be attached to a second polypeptide to provide for purification or detection of the second polypeptide or provide sites for attachment of the second polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract, protein A (Nilsson et al., EMBO J. 4:1075, 1985; Nilsson et al., Methods Enzymol. 198:3, 1991), glutathione S transferase (Smith and Johnson, Gene 67:31, 1988), Glu-Glu affinity tag (Grussenmeyer et al., Proc. Natl. Acad. Sci. USA 82:7952-4, 1985), substance P, Flag™ peptide (Hopp et al., Biotechnology 6:1204-10, 1988), streptavidin binding peptide, or other antigenic epitope or binding domain. See, in general, Ford et al., Protein Expression and Purification 2: 95-107, 1991. DNAs

encoding affinity tags are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, NJ).

The term "allelic variant" is used herein to denote any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene.

The terms "amino-terminal" and "carboxyl-terminal" are used herein to denote positions within polypeptides. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a polypeptide is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete polypeptide.

The term "complement/anti-complement pair" denotes non-identical moieties that form a non-covalently associated, stable pair under appropriate conditions. For instance, biotin and avidin (or streptavidin) are prototypical members of a complement/anti-complement pair. Other exemplary complement/anti-complement pairs include receptor/ligand pairs, antibody/antigen (or hapten or epitope) pairs, sense/antisense polynucleotide pairs, and the like. Where subsequent dissociation of the complement/anti-complement pair is desirable, the complement/anti-complement pair preferably has a binding affinity of $<10^9\ M^{-1}$.

The term "complements of a polynucleotide molecule" is a polynucleotide molecule having a complementary base sequence and reverse orientation as compared to a reference sequence. For example, the sequence 5' ATGCACGGG 3' is complementary to 5' CCCGTGCAT 3'.

The term "contig" denotes a polynucleotide that has a contiguous stretch of identical or complementary sequence to another polynucleotide. Contiguous sequences are said to "overlap" a given stretch of polynucleotide sequence either in their entirety or along a partial stretch of the polynucleotide. For example, representative contigs to the polynucleotide sequence 5'-ATGGCTTAGCTT-3' are 5'-TAGCTTgagtct-3' and 3'-gtcgacTACCGA-5'.

The term "degenerate nucleotide sequence" denotes a sequence of nucleotides that includes one or more degenerate codons (as compared to a reference polynucleotide molecule that encodes a polypeptide). Degenerate codons contain

different triplets of nucleotides, but encode the same amino acid residue (i.e., GAU and GAC triplets each encode Asp).

The term "expression vector" is used to denote a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that provide for its transcription. Such additional segments include promoter and terminator sequences, and may also include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both.

The term "isolated", when applied to a polynucleotide, denotes that the polynucleotide has been removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes with which they are ordinarily associated, but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators. The identification of associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, Nature 316:774-78, 1985).

An "isolated" polypeptide or protein is a polypeptide or protein that is found in a condition other than its native environment, such as apart from blood and animal tissue. In a preferred form, the isolated polypeptide is substantially free of other polypeptides, particularly other polypeptides of animal origin. It is preferred to provide the polypeptides in a highly purified form, i.e. greater than 95% pure, more preferably greater than 99% pure. When used in this context, the term "isolated" does not exclude the presence of the same polypeptide in alternative physical forms, such as dimers or alternatively glycosylated or derivatized forms.

The term "operably linked", when referring to DNA segments, indicates that the segments are arranged so that they function in concert for their intended purposes, e.g., transcription initiates in the promoter and proceeds through the coding segment to the terminator.

The term "ortholog" denotes a polypeptide or protein obtained from one species that is the functional counterpart of a polypeptide or protein from a different species. Sequence differences among orthologs are the result of speciation.

The term "ortholog" denotes a polypeptide or protein obtained from one species that is the functional counterpart of a polypeptide or protein from a different species. Sequence differences among orthologs are the result of speciation.

A "polynucleotide" is a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and may be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic molecules. Sizes of polynucleotides are expressed as base pairs (abbreviated "bp"), nucleotides ("nt"), or kilobases ("kb"). Where the context allows, the latter two terms may describe polynucleotides that are single-stranded or double-stranded. When the term is applied to double-stranded molecules it is used to denote overall length and will be understood to be equivalent to the term "base pairs". It will be recognized by those skilled in the art that the two strands of a double-stranded polynucleotide may differ slightly in length and that the ends thereof may be staggered as a result of enzymatic cleavage; thus all nucleotides within a double-stranded polynucleotide molecule may not be paired. Such unpaired ends will in general not exceed 20 nt in length.

A "polypeptide" is a polymer of amino acid residues joined by peptide bonds, whether produced naturally or synthetically. Polypeptides of less than about 10 amino acid residues are commonly referred to as "peptides".

The term "promoter" is used herein for its art-recognized meaning to denote a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of transcription. Promoter sequences are commonly, but not always, found in the 5' non-coding regions of genes.

A "protein" is a macromolecule comprising one or more polypeptide chains. A protein may also comprise non-peptidic components, such as carbohydrate groups. Carbohydrates and other non-peptidic substituents may be added to a protein by the cell in which the protein is produced, and will vary with the type of cell. Proteins are defined herein in terms of their amino acid backbone structures; substituents such as carbohydrate groups are generally not specified, but may be present nonetheless.

The term "receptor" denotes a cell-associated protein that binds to a bioactive molecule (i.e., a ligand) and mediates the effect of the ligand on the cell. Membrane-bound receptors are characterized by a multi-peptide structure comprising an extracellular ligand-binding domain and an intracellular effector domain that is typically involved in signal transduction. Binding of ligand to receptor results in a conformational change in the receptor that causes an interaction between the effector domain and other molecule(s) in the cell. This interaction in turn leads to an alteration in the metabolism of the cell. Metabolic events that are linked to receptor-ligand interactions include gene transcription, phosphorylation, dephosphorylation, increases in cyclic AMP production, mobilization of cellular calcium, mobilization of membrane

lipids, cell adhesion, hydrolysis of inositol lipids and hydrolysis of phospholipids. In general, receptors can be membrane bound, cytosolic or nuclear; monomeric (e.g., thyroid stimulating hormone receptor, beta-adrenergic receptor) or multimeric (e.g., PDGF receptor, growth hormone receptor, IL-3 receptor, GM-CSF receptor, G-CSF receptor, erythropoietin receptor and IL-6 receptor).

The term "secretory signal sequence" denotes a DNA sequence that encodes a polypeptide (a "secretory peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized. The larger polypeptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway.

The term "splice variant" is used herein to denote alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Molecular weights and lengths of polymers determined by imprecise analytical methods (e.g., gel electrophoresis) will be understood to be approximate values. When such a value is expressed as "about" X or "approximately" X, the stated value of X will be understood to be accurate to ±10%.

All references cited herein are incorporated by reference in their entirety.

The present invention is based in part upon the discovery of a novel DNA sequence that encodes a fibroblast growth factor (FGF) homolog polypeptide having approximately 54% homology to FGF-10 (Emoto et al., J. Biol. Chem. 272:23191-23194, 1997) and KGF-2 (PCT Patent Application WO 96/25422). The FGF homolog polypeptide has been designated zFGF-10.

The novel zFGF-10 polypeptides of the present invention contain a motif known to occur in all known members of the FGF family, and is unique to these proteins. The zFGF-10 homolog polypeptide encoded by DNA contains a motif of the formula: CXFXE, wherein X is any amino acid and X{} is the number of X amino acids greater than one (SEQ ID NO: 14). This motif is highly conserved in all members of the FGF family, and a consensus amino acid sequence of the domain includes, for example, human myocyte-activating factor (FGF-10; HSU76381, GENBANK identifier, NCBI, Bethesda, MD ;SEQ ID NO: 4), human fibroblast growth factor homologous factor 4 (FHF-4; Smallwood et al., 1996, *ibid.*; SEQ ID NO: 12), human

fibroblast growth factor homologous factor 3 (FHF-3; Smallwood et al., 1996, *ibid.*; SEQ ID NO: 10), human FGF-4 (Basilico et al., Adv. Cancer Res. 59:115-165,1992; SEQ ID NO: 9), human FGF-6 (Basilico et al., 1992, *ibid.*), human FGF-2 (basic; Basilico et al., 1992, *ibid.*), human FGF-1 (acidic; Basilico et al., 1992, *ibid.*), human keratinocyte growth factor precursor (FGF-7; Basilico et al., 1992, *ibid.*; SEQ ID NO: 5), human FGF-5 (Basilico et al., 1992, *ibid.*; SEQ ID NO: 11), human FGF-9 (Miyamoto et al., Mol. Cell. Biol. 13:4251-4259, 1993; SEQ ID NO: 7), human FGF-3 (Basilico et al., 1992, *ibid.*; SEQ ID NO: 8), human FGF-16 (Miyake et al., Biochem. Biophys. Res. Commun. 243(1):148-152(1998); SEQ ID NO: 6) and human FGF-12 (Kok et al., Biochem. Biophys. Res. Commun. 255(3):717-721, 1999; SEQ ID NO: 13).

The genomic DNA sequence as shown in SEQ ID NO. 1, consists of three exons separated by two introns. All the known FGF family members are spliced to form single polypeptides. The deduced amino acid sequence is shown in SEQ ID NO: 2. Analysis of the DNA encoding a zFGF-10 polypeptide (SEQ ID NO: 1) revealed three exons when spliced together formed an open reading frame encoding 170 amino acids (SEQ ID NO: 2) comprising a mature polypeptide of 153 amino acids (residue 18 to residue 170 of SEQ ID NO: 2) with a secretory signal sequence of 17 amino acids (residue 1 to 17 of SEQ ID NO: 2). Multiple alignment of zFGF-10 with other known FGFs revealed a block of high percent identity corresponding to amino acid residue 72 to 130 of SEQ ID NO: 2. The FGF family motif, as shown in SEQ ID NO: 14, corresponds to amino acid residues 113 (Cys) to 117 (Glu) of SEQ ID NO: 2. Several of the members of the FGF family do not have signal sequences.

Members of the FGF family are characterized by heparin binding domains. A putative heparin-binding domain for zFGF-10 has been identified in the region of amino acid residue 41 to amino acid residue 44 of SEQ ID NO: 2. It is postulated that receptor-mediated signaling is initiated upon binding of FGF ligand complexed with cell-surface heparin sulfate proteoglycans.

Based on homology alignments with FGF-1 and FGF-2 crystal structures (Eriksson et al., Prot. Sci. 2:1274, 1993), secondary structure predictions for beta strand structure of zFGF-10 includes the following regions of amino acid residues: strand 2—-51 (Phe)-57 (Pro); strand 3—59 (Gly)-64 (Thr); strand 4—72 (Ser)-78 (Ser); strand 5—82 (Gly)-89 (Val); strand 6—93 (Phe)-98 (Asn); strand 7—100 (Arg)-106 (Ser); strand 8—113 (Cys)-120 (Glu); strand 9—123 (Gly)-130 (Gln); strand 10—139 (Met)-144 (Asp); strand 11 148 (Gly)-152 (Gly) and strand 12—162 (Ala)-168 (Leu), as shown in SEQ ID NO: 2. Amino acids critical for zFGF-10 binding to receptors can be identified by site-directed mutagenesis of the entire zFGF-10 polypeptide. More specifically, they can be identified using site-directed mutagenesis of amino acids in the

zFGF-10 polypeptide which correspond to amino acid residues in acidic FGF (FGF1) and basic FGF (FGF2) which have been identified as critical for binding to their respective receptors (Blaber et al., Biochem. 35:2086-2094, 1996). In zFGF-10 hydrophobic residues buried within the core of the protein will be relatively intolerant of substitution, particularly polar or charged residues. Residues critical to the beta-trefoil fold of the zFGF-10 include residues 53 (Leu), 61 (Val), 74 (Leu), 76 (Ile), 84 (Val), 86 (Ile), 95 (Val), 103 (Leu), 115 (Phe), 119 (Ile), 127 (Tyr), 141 (Leu) and 164 (Phe). One skilled in the art will recognize that other members, in whole or in part, of the FGF family may have structural or biochemical similarities to zFGF-10. Therefore, amino acid residues from another FGF family member can be used for substitutions at corresponding positions in zFGF-10 given the limitations disclosed herein. Corresponding amino acids can be identified using a multiple alignment.

Those skilled in the art will recognize that predicted domain boundaries are somewhat imprecise and may vary by up to ± 3 amino acid residues.

Polypeptides of the present invention comprise at least 6, at least 9, or at least 15 contiguous amino acid residues of SEQ ID NO:2. Within certain embodiments of the invention, the polypeptides comprise 20, 30, 40, 50, 100, or more contiguous residues of SEQ ID NO:2, up to the entire predicted mature polypeptide (residues 18 to 170 of SEQ ID NO:2) or the primary translation product (residues 1 to 170 of SEQ ID NO:2). As disclosed in more detail below, these polypeptides can further comprise additional, non-zFGF-10, polypeptide sequence(s).

Within the polypeptides of the present invention are polypeptides that comprise an epitope-bearing portion of a protein as shown in SEQ ID NO:2. An "epitope" is a region of a protein to which an antibody can bind. See, for example, Geysen et al., *Proc. Natl. Acad. Sci. USA* 81:3998-4002, 1984. Epitopes can be linear or conformational, the latter being composed of discontinuous regions of the protein that form an epitope upon folding of the protein. Linear epitopes are generally at least 6 amino acid residues in length. Relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, Sutcliffe et al., *Science* 219:660-666, 1983. Antibodies that recognize short, linear epitopes are particularly useful in analytic and diagnostic applications that employ denatured protein, such as Western blotting (Tobin, *Proc. Natl. Acad. Sci. USA* 76:4350-4356, 1979), or in the analysis of fixed cells or tissue samples. Antibodies to linear epitopes are also useful for detecting fragments of zFGF-10, such as might occur in body fluids or cell culture media.

Antigenic, epitope-bearing polypeptides of the present invention are useful for raising antibodies, including monoclonal antibodies, that specifically bind to

a zFGF-10 protein. Antigenic, epitope-bearing polypeptides contain a sequence of at least six, or at least nine, and in some embodiments from 15 to about 30 contiguous amino acid residues of a zFGF-10 protein (e.g., SEQ ID NO:2). Polypeptides comprising a larger portion of a zFGF-10 protein, i.e. from 30 to 50 residues up to the entire sequence, are included. It is preferred that the amino acid sequence of the epitope-bearing polypeptide is selected to provide substantial solubility in aqueous solvents, that is the sequence includes relatively hydrophilic residues, and hydrophobic residues are substantially avoided. Specific, useful polypeptides in this regard include those comprising residues 116-121, 130-135, 144-149, 153-158, and 152-157 of SEQ ID NO:2.

Polypeptides of the present invention can be prepared with one or more amino acid substitutions, deletions or additions as compared to SEQ ID NO:2. These changes are preferably of a minor nature, that is conservative amino acid substitutions and other changes that do not significantly affect the folding or activity of the protein or polypeptide as described herein. These changes include amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, an amino or carboxyl-terminal cysteine residue to facilitate subsequent linking to maleimide-activated keyhole limpet hemocyanin, a small linker peptide of up to about 20-25 residues, or an extension that facilitates purification (an affinity tag) as disclosed above. Two or more affinity tags may be used in combination. Polypeptides comprising affinity tags can further comprise a polypeptide linker and/or a proteolytic cleavage site between the zFGF-10 polypeptide and the affinity tag. Preferred cleavage sites include thrombin cleavage sites and factor Xa cleavage sites.

The present invention further provides a variety of other polypeptide fusions. For example, a zFGF-10 polypeptide can be prepared as a fusion to a dimerizing protein as disclosed in U.S. Patents Nos. 5,155,027 and 5,567,584. Preferred dimerizing proteins in this regard include immunoglobulin constant region domains. Immunoglobulin-zFGF-10 polypeptide fusions can be expressed in genetically engineered cells to produce a variety of multimeric zFGF-10 analogs. In addition, a zFGF-10 polypeptide can be joined to another bioactive molecule, such as a cytokine, to provide a multi-functional molecule. One or more helices of a zFGF-10 polypeptide can be joined to another cytokine to enhance or otherwise modify its biological properties. Auxiliary domains can be fused to zFGF-10 polypeptides to target them to specific cells, tissues, or macromolecules (e.g., collagen). For example, a zFGF-10 polypeptide or protein can be targeted to a predetermined cell type by fusing a zFGF-10 polypeptide to a ligand that specifically binds to a receptor on the surface of the target cell. In this way, polypeptides and proteins can be targeted for therapeutic or

diagnostic purposes. A zFGF-10 polypeptide can be fused to two or more moieties, such as an affinity tag for purification and a targeting domain. Polypeptide fusions can also comprise one or more cleavage sites, particularly between domains. See, Tuan et al., *Connective Tissue Research* 34:1-9, 1996. Thus, the present invention includes fusion proteins of at least two moieties, wherein either the first or second moiety comprises a zFGF10 polypeptide or fragment thereof, and the other moieties are derived from, but are not limited to, an affinity tag, a corresponding region from another FGF family protein, a cytokine, a targeting molecule.

Polypeptide fusions of the present invention will generally contain not more than about 1,500 amino acid residues, preferably not more than about 1,200 residues, more preferably not more than about 1,000 residues, and will in many cases be considerably smaller. For example, a zFGF-10 polypeptide of 337 residues (residues 18-354 of SEQ ID NO:2) can be fused to *E. coli* β-galactosidase (1,021 residues; see Casadaban et al., *J. Bacteriol.* 143:971-980, 1980), a 10-residue spacer, and a 4-residue factor Xa cleavage site to yield a polypeptide of 1,188 residues. In a second example, residues 18-170 SEQ ID NO:2 can be fused to maltose binding protein (approximately 370 residues), a 4-residue cleavage site, and a 6-residue polyhistidine tag.

As disclosed above, the polypeptides of the present invention comprise at least 6 contiguous residues of SEQ ID NO:2. These polypeptides may further comprise additional residues as shown in SEQ ID NO:2, a variant of SEQ ID NO:2, or another protein as disclosed herein. When variants of SEQ ID NO:2 are employed, the resulting polypeptide will be at least 90%, or at least 95% to 99% identical to the corresponding region of SEQ ID NO:2. Percent sequence identity is determined by conventional methods. See, for example, Altschul et al., *Bull. Math. Bio.* 48:603-616, 1986, and Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915-10919, 1992. Briefly, two amino acid sequences are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "BLOSUM62" scoring matrix of Henikoff and Henikoff (*ibid.*) as shown in Table 1 (amino acids are indicated by the standard one-letter codes). The percent identity is then calculated as:

$$\frac{\text{Total number of identical matches}}{\left[\begin{array}{l}\text{length of the longer sequence plus the} \\ \text{number of gaps introduced into the longer} \\ \text{sequence in order to align the two} \\ \text{sequences}\end{array}\right]} \times 100$$

Table 1

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

The level of identity between amino acid sequences can be determined using the "FASTA" similarity search algorithm disclosed by Pearson and Lipman (*Proc. Natl. Acad. Sci. USA* 85:2444, 1988) and by Pearson (*Meth. Enzymol.* 183:63, 1990). Briefly, FASTA first characterizes sequence similarity by identifying regions shared by the query sequence (e.g., SEQ ID NO:2) and a test sequence that have either the highest density of identities (if the ktup variable is 1) or pairs of identities (if ktup=2), without considering conservative amino acid substitutions, insertions, or deletions. The ten regions with the highest density of identities are then rescored by comparing the similarity of all paired amino acids using an amino acid substitution matrix, and the ends of the regions are "trimmed" to include only those residues that contribute to the highest score. If there are several regions with scores greater than the "cutoff" value (calculated by a predetermined formula based upon the length of the sequence and the ktup value), then the trimmed initial regions are examined to determine whether the regions can be joined to form an approximate alignment with gaps. Finally, the highest scoring regions of the two amino acid sequences are aligned using a modification of the Needleman-Wunsch-Sellers algorithm (Needleman and Wunsch, *J. Mol. Biol.* 48:444, 1970; Sellers, *SIAM J. Appl. Math.* 26:787, 1974), which allows for amino acid insertions and deletions. Preferred parameters for FASTA analysis are: ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=BLOSUM62. These parameters can be introduced into a FASTA program by modifying the scoring matrix file ("SMATRIX"), as explained in Appendix 2 of Pearson, 1990 (*ibid.*).

FASTA can also be used to determine the sequence identity of nucleic acid molecules using a ratio as disclosed above. For nucleotide sequence comparisons, the ktup value can range between one to six, preferably from three to six, most preferably three, with other parameters set as default.

The present invention includes polypeptides having one or more conservative amino acid changes as compared with the amino acid sequence of SEQ ID NO:2. The BLOSUM62 matrix (Table 1) is an amino acid substitution matrix derived from about 2,000 local multiple alignments of protein sequence segments, representing highly conserved regions of more than 500 groups of related proteins (Henikoff and Henikoff, *ibid.*). Thus, the BLOSUM62 substitution frequencies can be used to define conservative amino acid substitutions that may be introduced into the amino acid sequences of the present invention. As used herein, the term "conservative amino acid substitution" refers to a substitution represented by a BLOSUM62 value of greater than -1. For example, an amino acid substitution is conservative if the substitution is

characterized by a BLOSUM62 value of 0, 1, 2, or 3. Preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least one 1 (e.g., 1, 2 or 3), while more preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least 2 (e.g., 2 or 3).

The proteins of the present invention can also comprise non-naturally occuring amino acid residues. Non-naturally occuring amino acids include, without limitation, *trans*-3-methylproline, 2,4-methanoproline, *cis*-4-hydroxyproline, *trans*-4-hydroxyproline, *N*-methylglycine, *allo*-threonine, methylthreonine, hydroxyethylcysteine, hydroxyethylhomocysteine, nitroglutamine, homoglutamine, pipecolic acid, *tert*-leucine, norvaline, 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occuring amino acid residues into proteins. For example, an *in vitro* system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is carried out in a cell-free system comprising an *E. coli* S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson et al., *J. Am. Chem. Soc.* 113:2722, 1991; Ellman et al., *Methods Enzymol.* 202:301, 1991; Chung et al., *Science* 259:806-809, 1993; and Chung et al., *Proc. Natl. Acad. Sci. USA* 90:10145-10149, 1993). In a second method, translation is carried out in Xenopus oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., *J. Biol. Chem.* 271:19991-19998, 1996). Within a third method, *E. coli* cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occuring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occuring amino acid is incorporated into the protein in place of its natural counterpart. See, Koide et al., *Biochem.* 33:7470-7476, 1994. Naturally occuring amino acid residues can be converted to non-naturally occuring species by *in vitro* chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, *Protein Sci.* 2:395-403, 1993).

Amino acid sequence changes are made in zFGF-10 polypeptides so as to minimize disruption of higher order structure essential to biological activity as disclosed previously. Amino acid residues that are within regions or domains that are critical to maintaining structural integrity can be determined. Within these regions one

can identify specific residues that will be more or less tolerant of change and maintain the overall tertiary structure of the molecule. Methods for analyzing sequence structure include, but are not limited to, alignment of multiple sequences with high amino acid or nucleotide identity, secondary structure propensities, binary patterns, complementary packing, and buried polar interactions (Barton, *Current Opin. Struct. Biol.* 5:372-376, 1995 and Cordes et al., *Current Opin. Struct. Biol.* 6:3-10, 1996). In general, determination of structure will be accompanied by evaluation of activity of modified molecules. For example, changes in amino acid residues will be made so as not to disrupt the beta-trefoil fold structure of the protein family. The effects of amino acid sequence changes can be predicted by, for example, computer modeling using available software (e.g., the Insight II® viewer and homology modeling tools; MSI, San Diego, CA) or determined by analysis of crystal structure (see, e.g., Lapthorn et al, *Nature* 369:455-461, 1994; Lapthorn et al., *Nat. Struct. Biol.* 2:266-268, 1995). Protein folding can be measured by circular dichroism (CD). Measuring and comparing the CD spectra generated by a modified molecule and standard molecule are routine in the art (Johnson, *Proteins* 7:205-214, 1990). Crystallography is another well known and accepted method for analyzing folding and structure. Nuclear magnetic resonance (NMR), digestive peptide mapping and epitope mapping are other known methods for analyzing folding and structural similarities between proteins and polypeptides (Schaanan et al., *Science* 257:961-964, 1992). Mass spectrometry and chemical modification using reduction and alkylation can be used to identify cysteine residues that are associated with disulfide bonds or are free of such associations (Bean et al., *Anal. Biochem.* 201:216-226, 1992; Gray, *Protein Sci.* 2:1732-1748, 1993; and Patterson et al., *Anal. Chem.* 66:3727-3732, 1994). Alterations in disulfide bonding will be expected to affect protein folding. These techniques can be employed individually or in combination to analyze and compare the structural features that affect folding of a variant protein or polypeptide to a standard molecule to determine whether such modifications would be significant.

Essential amino acids in the polypeptides of the present invention can be identified experimentally according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* 244, 1081-1085, 1989; Bass et al., *Proc. Natl. Acad. Sci. USA* 88:4498-4502, 1991). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for biological activity as disclosed below to identify amino acid residues that are critical to the activity of the molecule.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (*Science* 241:53-57, 1988) or Bowie and Sauer (*Proc. Natl. Acad. Sci. USA* 86:2152-2156, 1989). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., *Biochem.* 30:10832-10837, 1991; Ladner et al., U.S. Patent No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., *Gene* 46:145, 1986; Ner et al., *DNA* 7:127, 1988).

Variants of the disclosed zFGF-10 DNA and polypeptide sequences can be generated through DNA shuffling as disclosed by Stemmer, *Nature* 370:389-391, 1994 and Stemmer, *Proc. Natl. Acad. Sci. USA* 91:10747-10751, 1994. Briefly, variant genes are generated by *in vitro* homologous recombination by random fragmentation of a parent gene followed by reassembly using PCR, resulting in randomly introduced point mutations. This technique can be modified by using a family of parent genes, such as allelic variants or genes from different species, to introduce additional variability into the process. Selection or screening for the desired activity, followed by additional iterations of mutagenesis and assay provides for rapid "evolution" of sequences by selecting for desirable mutations while simultaneously selecting against detrimental changes.

In many cases, the structure of the final polypeptide product will result from processing of the nascent polypeptide chain by the host cell, thus the final sequence of a zFGF-10 polypeptide produced by a host cell will not always correspond to the full sequence encoded by the expressed polynucleotide. For example, expressing the complete zFGF-10 sequence in a cultured mammalian cell is expected to result in removal of at least the secretory peptide, while the same polypeptide produced in a prokaryotic host would not be expected to be cleaved. Differential processing of individual chains may result in heterogeneity of expressed polypeptides.

SEQ ID NO: 3 is a degenerate polynucleotide sequence that encompasses all polynucleotides that could encode the zFGF-10 polypeptide of SEQ ID NO: 2 (amino acids 1 or 18 to 170). Thus, zFGF-10 polypeptide-encoding polynucleotides ranging from nucleotide 1 or 51 to nucleotide 510 of SEQ ID NO: 3 are contemplated by the present invention. Also contemplated by the present invention are fragments and fusions as described above with respect to SEQ ID NO: 1, which are

formed from analogous regions of SEQ ID NO: 3, wherein nucleotides 1 or 51 to 510 of SEQ ID NO: 3 correspond to nucleotides 1 or 51 to 510 of SEQ ID NO: 1, for the encoding a mature zFGF-10 molecule.

The symbols in SEQ ID NO: 3 are summarized in Table 1 below.

TABLE 1

Nucleotide	Resolutions	Complement	Resolutions
A	A	T	T
C	C	G	G
G	G	C	C
T	T	A	A
R	A|G	Y	C|T
Y	C|T	R	A|G
M	A|C	K	G|T
K	G|T	M	A|C
S	C|G	S	C|G
C|G	A|T	W	A|T
H	A|C|T	D	A|G|T
B	C|G|T	V	A|C|G
V	A|C|G	B	C|G|T
D	A|G|T	H	A|C|T
N	A|C|G|T	N	A|C|G|T

The degenerate codons used in SEQ ID NO: 3, encompassing all possible codons for a given amino acid, are set forth in Table 2.

TABLE 2

Amino Acid	Letter	Codons	Degenerate Codon
Cys	C	TGC TGT	TGY
Ser	S	AGC AGT TCA TCC TCG TCT	WSN
Thr	T	ACA ACC ACG ACT	ACN
Pro	P	CCA CCC CCG CCT	CCN
Ala	A	GCA GCC GCG GCT	GCN
Gly	G	GGA GGC GGG GGT	GGN
Asn	N	AAC AAT	AAY
Asp	D	GAC GAT	GAY
Glu	E	GAA GAG	GAR
Gln	Q	CAA CAG	CAR
His	H	CAC CAT	CAY
Arg	R	AGA AGG CGA CGC CGG CGT	MGN
Lys	K	AAA AAG	AAR
Met	M	ATG	ATG
Ile	I	ATA ATC ATT	ATH
Leu	L	CTA CTC CTG CTT TTA TTG	YTN
Val	V	GTA GTC GTG GTT	GTN
Phe	F	TTC TTT	TTY
Tyr	Y	TAC TAT	TAY
Trp	W	TGG	TGG
Ter	.	TAA TAG TGA	TRR
Asn|Asp	B		RAY
Glu|Gln	Z		SAR
Any	X		NNN
Gap	-	---	

One of ordinary skill in the art will appreciate that some ambiguity is introduced in determining a degenerate codon, representative of all possible codons encoding each amino acid. For example, the degenerate codon for serine (WSN) can, in some circumstances, encode arginine (AGR), and the degenerate codon for arginine (MGN) can, in some circumstances, encode serine (AGY). A similar relationship exists

between codons encoding phenylalanine and leucine. Thus, some polynucleotides encompassed by the degenerate sequence may have some variant amino acids, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequence of SEQ ID NO: 2. Variant sequences can be readily tested for functionality as described herein.

One of ordinary skill in the art will also appreciate that different species can exhibit "preferential codon usage." In general, see, Grantham, et al., Nuc. Acids Res. 8:1893-912, 1980; Haas, et al. Curr. Biol. 6:315-24, 1996; Wain-Hobson, et al., Gene 13:355-64, 1981; Grosjean and Fiers, Gene 18:199-209, 1982; Holm, Nuc. Acids Res. 14:3075-87, 1986; Ikemura, J. Mol. Biol. 158:573-97, 1982. As used herein, the term "preferential codon usage" or "preferential codons" is a term of art referring to protein translation codons that are most frequently used in cells of a certain species, thus favoring one or a few representatives of the possible codons encoding each amino acid (See Table 2). For example, the amino acid Threonine (Thr) may be encoded by ACA, ACC, ACG, or ACT, but in mammalian cells ACC is the most commonly used codon; in other species, for example, insect cells, yeast, viruses or bacteria, different Thr codons may be preferential. Preferential codons for a particular species can be introduced into the polynucleotides of the present invention by a variety of methods known in the art. Introduction of preferential codon sequences into recombinant DNA can, for example, enhance production of the protein by making protein translation more efficient within a particular cell type or species. Therefore, the degenerate codon sequence disclosed in SEQ ID NO: 3 serves as a template for optimizing expression of polynucleotides in various cell types and species commonly used in the art and disclosed herein. Sequences containing preferential codons can be tested and optimized for expression in various species, and tested for functionality as disclosed herein.

The highly conserved amino acids in zFGF-10 can be used as a tool to identify new family members. To identify new family members in EST databases, the conserved CXFXE motif (SEQ ID NO: 14) can be used. In another method using polynucleotide probes and hybridization methods, RNA obtained from a variety of tissue sources can be used to generate cDNA libraries and probe these libraries for new family members. In particular, reverse transcription-polymerase chain reaction (RT-PCR) can be used to amplify sequences encoding highly degenerate DNA primers designed from the sequences corresponding to amino acid residue 113 (Cys) to amino acid residue 117 (Glu) of SEQ ID NO: 2.

The zFGF-10 gene has been mapped chromosome 19 (The Genetic Location Database, University of Southhampton, UK). Thus, the present invention

provides methods for using zFGF-10 polynucleotides and polypeptides to identify chromosomal disorders associated with abnormal expression of the zFGF-10 protein. zFGF-10 maps to a chromosomal locus that has been associated with psoriasis susceptibility (Lee et al., *Am. J. Human Genet.* 67:1020-1024, 2000). In view of the homology of zFGF-10 with KGF, and that the larger family of fibroblast growth factor proteins is known to have activity on keratinocytes, stromal cells and other cells of mesenchymal ontogeny, zFGF-10 provides a good candidate marker for this disease.

Detectable chromosomal mutations at the zFGF-10 gene locus include, but are not limited to, aneuploidy, gene copy number changes, insertions, deletions, translocations, restriction site changes and rearrangements. Such aberrations can be identified by employing molecular genetic techniques, such as restriction fragment length polymorphism (RFLP) analysis, short tandem repeat (STR) analysis employing PCR techniques, and other genetic linkage analysis techniques known in the art (Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and Ausubel et al., eds., Current Protocols in Molecular Biology, John Wiley and Sons, Inc., NY, 1987; A.J. Marian, Chest 108:255-65, 1995). Analyses of DNA samples can detect deletions and insertions by changes in size in amplified DNA products by comparing a sample DNA to a normal zFGF-10 DNA standard. Mismatches in duplex DNA can be detected by RNase digestion or differences in melting temperature. Other methods for detecting differences in sequences include changes in electrophoretic motility, Southern analysis, and direct DNA sequencing. Recently, techniques for accessing genetic information with high-density arrays have been available (Chee et al., Science 274:610-614, 1996), and can analyze large fragments of genomic DNA with high resolution.

Analysis of chromosomal DNA using the zFGF-10 polynucleotide sequence is useful for correlating disease with mutations localized to the chromosome where the zFGF-10 gene resides. Studies of the DNA sequences, cDNA and/or genomic DNA, of individuals presenting disease that correlates with a mutation in the sequence of the zFGF-10 gene, wherein such mutation is not present in normal individuals, can provide strong evidence for the mutation as causative factor of the disease. In one embodiment, the methods of the present invention provide a method of detecting a zFGF-10 chromosomal abnormality in sample from an individual comprising: (a) obtaining a zFGF-10 RNA from the sample; (b) generating zFGF-10 cDNA by polymerase chain reaction; and (c) comparing the nucleic acid sequence of the zFGF-10 cDNA to the nucleic acid sequence as shown in SEQ ID NO: 1. In further embodiments, the difference between the sequence of the zFGF-10 cDNA or zFGF-10

gene in the sample and the zFGF-10 sequence as shown in SEQ ID NO: 1 is indicative of a zFGF-10 chromosomal mutation.

Aberrations associated with a *ZFGF10* locus can be detected using nucleic acid molecules of the present invention by employing molecular genetic techniques, such as restriction fragment length polymorphism analysis, short tandem repeat analysis employing PCR techniques, amplification-refractory mutation system analysis, single-strand conformation polymorphism detection, RNase cleavage methods, denaturing gradient gel electrophoresis, fluorescence-assisted mismatch analysis, and other genetic analysis techniques known in the art (see, for example, Mathew (ed.), *Protocols in Human Molecular Genetics* (Humana Press, Inc. 1991), Marian, *Chest 108*:255 (1995), Coleman and Tsongalis, *Molecular Diagnostics* (Human Press, Inc. 1996), Elles (ed.) *Molecular Diagnosis of Genetic Diseases* (Humana Press, Inc. 1996), Landegren (ed.), *Laboratory Protocols for Mutation Detection* (Oxford University Press 1996), Birren *et al.* (eds.), *Genome Analysis, Vol. 2: Detecting Genes* (Cold Spring Harbor Laboratory Press 1998), Dracopoli *et al.* (eds.), *Current Protocols in Human Genetics* (John Wiley & Sons 1998), and Richards and Ward, "Molecular Diagnostic Testing," in *Principles of Molecular Medicine*, pages 83-88 (Humana Press, Inc. 1998)).

The protein truncation test is also useful for detecting the inactivation of a gene in which translation-terminating mutations produce only portions of the encoded protein (see, for example, Stoppa-Lyonnet *et al., Blood 91*:3920 (1998)). According to this approach, RNA is isolated from a biological sample, and used to synthesize cDNA. PCR is then used to amplify the *ZFGF10* target sequence and to introduce an RNA polymerase promoter, a translation initiation sequence, and an in-frame ATG triplet. PCR products are transcribed using an RNA polymerase, and the transcripts are translated *in vitro* with a T7-coupled reticulocyte lysate system. The translation products are then fractionated by SDS-PAGE to determine the lengths of the translation products. The protein truncation test is described, for example, by Dracopoli *et al.* (eds.), *Current Protocols in Human Genetics*, pages 9.11.1 - 9.11.18 (John Wiley & Sons 1998).

The present invention also contemplates kits for performing a diagnostic assay for *ZFGF10* gene expression or to analyze the *ZFGF10* locus of a subject. Such kits comprise nucleic acid probes, such as double-stranded nucleic acid molecules comprising the nucleotide sequence of SEQ ID NOS:1 or 9, or a fragment thereof, as well as single-stranded nucleic acid molecules having the complement of the nucleotide sequence of SEQ ID NOS:1 or 9, or a fragment thereof. Probe molecules may be DNA,

RNA, oligonucleotides, and the like. Kits may comprise nucleic acid primers for performing PCR.

Such a kit can contain all the necessary elements to perform a nucleic acid diagnostic assay described above. A kit will comprise at least one container comprising a ZFGF10 probe or primer. The kit may also comprise a second container comprising one or more reagents capable of indicating the presence of ZFGF10 sequences. Examples of such indicator reagents include detectable labels such as radioactive labels, fluorochromes, chemiluminescent agents, and the like. A kit may also comprise a means for conveying to the user that the ZFGF10 probes and primers are used to detect *ZFGF10* gene expression. For example, written instructions may state that the enclosed nucleic acid molecules can be used to detect either a nucleic acid molecule that encodes ZFGF10, or a nucleic acid molecule having a nucleotide sequence that is complementary to a ZFGF10-encoding nucleotide sequence, or to analyze chromosomal sequences associated with the *ZFGF10* locus. The written material can be applied directly to a container, or the written material can be provided in the form of a packaging insert.

Within preferred embodiments of the invention, the isolated nucleic acid molecules can hybridize under stringent conditions to nucleic acid molecules having at least a portion of the nucleotide sequence of SEQ ID NOS:1 or 5 or to nucleic acid molecules having a nucleotide sequence complementary to those sequences. A pair of nucleic acid molecules, such as DNA-DNA, RNA-RNA and DNA-RNA, can hybridize if the nucleotide sequences have some degree of complementarity. Hybrids can tolerate mismatched base pairs in the double helix, but the stability of the hybrid is influenced by the degree of mismatch. The T_m of the mismatched hybrid decreases by 1°C for every 1-1.5% base pair mismatch. Varying the stringency of the hybridization conditions allows control over the degree of mismatch that will be present in the hybrid. The degree of stringency increases as the hybridization temperature increases and the ionic strength of the hybridization buffer decreases. Stringent hybridization conditions encompass temperatures of about 5-25°C below the T_m of the hybrid and a hybridization buffer having up to 1 M Na^+. Higher degrees of stringency at lower temperatures can be achieved with the addition of formamide which reduces the T_m of the hybrid about 1°C for each 1% formamide in the buffer solution. Generally, such stringent conditions include temperatures of 20-70°C and a hybridization buffer containing up to 6xSSC and 0-50% formamide. A higher degree of stringency can be achieved at temperatures of from 40-70°C with a hybridization buffer having up to 4xSSC and from 0-50% formamide. Highly stringent conditions typically encompass

temperatures of 42-70°C with a hybridization buffer having up to 1xSSC and 0-50% formamide. Different degrees of stringency can be used during hybridization and washing to achieve maximum specific binding to the target sequence. Typically, the washes following hybridization are performed at increasing degrees of stringency to remove non-hybridized polynucleotide probes from hybridized complexes.

The above conditions are meant to serve as a guide and it is well within the abilities of one skilled in the art to adapt these conditions for use with a particular polypeptide hybrid. The T_m for a specific target sequence is the temperature (under defined conditions) at which 50% of the target sequence will hybridize to a perfectly matched probe sequence. Those conditions which influence the T_m include, the size and base pair content of the polynucleotide probe, the ionic strength of the hybridization solution, and the presence of destabilizing agents in the hybridization solution. Numerous equations for calculating T_m are known in the art, and are specific for DNA, RNA and DNA-RNA hybrids and polynucleotide probe sequences of varying length (see, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor Press 1989); Ausubel et al., (eds.), Current Protocols in Molecular Biology (John Wiley and Sons, Inc. 1987); Berger and Kimmel (eds.), Guide to Molecular Cloning Techniques, (Academic Press, Inc. 1987); and Wetmur, Crit. Rev. Biochem. Mol. Biol. 26:227 (1990)). Sequence analysis software, such as OLIGO 6.0 (LSR; Long Lake, MN) and Primer Premier 4.0 (Premier Biosoft International; Palo Alto, CA), as well as sites on the Internet, are available tools for analyzing a given sequence and calculating T_m based on user defined criteria. Such programs can also analyze a given sequence under defined conditions and identify suitable probe sequences. Typically, hybridization of longer polynucleotide sequences, >50 base pairs, is performed at temperatures of about 20-25°C below the calculated T_m. For smaller probes, <50 base pairs, hybridization is typically carried out at the T_m or 5-10°C below. This allows for the maximum rate of hybridization for DNA-DNA and DNA-RNA hybrids.

The length of the polynucleotide sequence influences the rate and stability of hybrid formation. Smaller probe sequences, <50 base pairs, reach equilibrium with complementary sequences rapidly, but may form less stable hybrids. Incubation times of anywhere from minutes to hours can be used to achieve hybrid formation. Longer probe sequences come to equilibrium more slowly, but form more stable complexes even at lower temperatures. Incubations are allowed to proceed overnight or longer. Generally, incubations are carried out for a period equal to three times the calculated Cot time. Cot time, the time it takes for the polynucleotide

sequences to reassociate, can be calculated for a particular sequence by methods known in the art.

The base pair composition of polynucleotide sequence will effect the thermal stability of the hybrid complex, thereby influencing the choice of hybridization temperature and the ionic strength of the hybridization buffer. A-T pairs are less stable than G-C pairs in aqueous solutions containing sodium chloride. Therefore, the higher the G-C content, the more stable the hybrid. Even distribution of G and C residues within the sequence also contribute positively to hybrid stability. In addition, the base pair composition can be manipulated to alter the T_m of a given sequence. For example, 5-methyldeoxycytidine can be substituted for deoxycytidine and 5-bromodeoxuridine can be substituted for thymidine to increase the T_m, whereas 7-deazz-2'-deoxyguanosine can be substituted for guanosine to reduce dependence on T_m.

The ionic concentration of the hybridization buffer also affects the stability of the hybrid. Hybridization buffers generally contain blocking agents such as Denhardt's solution (Sigma Chemical Co., St. Louis, Mo.), denatured salmon sperm DNA, tRNA, milk powders (BLOTTO), heparin or SDS, and a Na^+ source, such as SSC (1x SSC: 0.15 M sodium chloride, 15 mM sodium citrate) or SSPE (1x SSPE: 1.8 M NaCl, 10 mM NaH_2PO_4, 1 mM EDTA, pH 7.7). By decreasing the ionic concentration of the buffer, the stability of the hybrid is increased. Typically, hybridization buffers contain from between 10 mM - 1 M Na^+. The addition of destabilizing or denaturing agents such as formamide, tetralkylammonium salts, guanidinium cations or thiocyanate cations to the hybridization solution will alter the T_m of a hybrid. Typically, formamide is used at a concentration of up to 50% to allow incubations to be carried out at more convenient and lower temperatures. Formamide also acts to reduce non-specific background when using RNA probes.

As previously noted, the isolated polynucleotides of the present invention include DNA and RNA. Methods for preparing DNA and RNA are well known in the art. In general, RNA is isolated from a tissue or cell that produces large amounts of zFGF-10 RNA. Such tissues and cells are identified by Northern blotting (Thomas, Proc. Natl. Acad. Sci. USA 77:5201, 1980), and include pancreas and prostate. Total RNA can be prepared using guanidinium isothiocyanate extraction followed by isolation by centrifugation in a CsCl gradient (Chirgwin et al., Biochemistry 18:52-94, 1979). Poly $(A)^+$ RNA is prepared from total RNA using the method of Aviv and Leder (Proc. Natl. Acad. Sci. USA 69:1408-1412, 1972). Complementary DNA (cDNA) is prepared from poly$(A)^+$ RNA using known methods.

Polynucleotides encoding zFGF-10 polypeptides are then identified and isolated by, for example, hybridization or PCR.

A full-length clone encoding zFGF-10 can be obtained by conventional cloning procedures. Complementary DNA (cDNA) clones are preferred, although for some applications (e.g., expression in transgenic animals) it may be preferable to use a genomic clone, or to modify a cDNA clone to include at least one genomic intron. Methods for preparing cDNA and genomic clones are well known and within the level of ordinary skill in the art, and include the use of the sequence disclosed herein, or parts thereof, for probing or priming a library. Expression libraries can be probed with antibodies to zFGF-10, receptor fragments, or other specific binding partners.

The present invention further provides counterpart polypeptides and polynucleotides from other species (orthologs). Of particular interest are zFGF-10 polypeptides from other mammalian species, including murine, rat, porcine, ovine, bovine, canine, feline, equine and other primate proteins.

Orthologs of the human proteins can be cloned using information and compositions provided by the present invention in combination with conventional cloning techniques. For example, a cDNA can be cloned using mRNA obtained from a tissue or cell type that expresses the protein. Suitable sources of mRNA can be identified by probing Northern blots with probes designed from the sequences disclosed herein. A library is then prepared from mRNA of a positive tissue or cell line. A zFGF-10-encoding cDNA can then be isolated by a variety of methods, such as by probing with a complete or partial human cDNA or with one or more sets of degenerate probes based on the disclosed sequences. A cDNA can also be cloned using the polymerase chain reaction, or PCR (Mullis, U.S. Patent 4,683,202), using primers designed from the sequences disclosed herein. Within an additional method, the cDNA library can be used to transform or transfect host cells, and expression of the cDNA of interest can be detected with an antibody to zFGF-10. Similar techniques can also be applied to the isolation of genomic clones.

Those skilled in the art will recognize that the sequences disclosed in SEQ ID NO: 1 and SEQ ID NO: 2 represent a single allele of the human zFGF-10 gene and polypeptide, respectively, and that allelic variation and alternative splicing are expected to occur. Allelic variants can be cloned by probing cDNA or genomic libraries from different individuals according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO: 1, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are

within the scope of the present invention, as are proteins which are allelic variants of SEQ ID NO: 2.

Mutagenesis methods as disclosed above can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides in host cells. Mutagenized DNA molecules that encode active polypeptides (e.g., cell proliferation) can be recovered from the host cells and rapidly sequenced using modern equipment. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide of interest, and can be applied to polypeptides of unknown structure.

Variants of the disclosed zFGF-10 DNA and polypeptide sequences can be generated through DNA shuffling as disclosed by Stemmer, Nature 370:389-91, 1994, Stemmer, Proc. Natl. Acad. Sci. USA 91:10747-51, 1994 and WIPO Publication WO 97/20078. Briefly, variant DNAs are generated by in vitro homologous recombination by random fragmentation of a parent DNA followed by reassembly using PCR, resulting in randomly introduced point mutations. This technique can be modified by using a family of parent DNAs, such as allelic variants or DNAs from different species, to introduce additional variability into the process. Selection or screening for the desired activity, followed by additional iterations of mutagenesis and assay provides for rapid "evolution" of sequences by selecting for desirable mutations while simultaneously selecting against detrimental changes.

Using the methods discussed above, one of ordinary skill in the art can identify and/or prepare a variety of polypeptides that are substantially homologous to residues 18 (Pro) to 170 (Ser) or residues 1 (Met) to 170 (Ser) of SEQ ID NO: 2, allelic variants thereof, or biologically active fragments thereof, and retain the proliferative properties of the wild-type protein. Such polypeptides may also include additional polypeptide segments as generally disclosed above.

The polypeptides of the present invention, including full-length proteins, fragments thereof and fusion proteins, can be produced in genetically engineered host cells according to conventional techniques. Suitable host cells are those cell types that can be transformed or transfected with exogenous DNA and grown in culture, and include bacteria, fungal cells, and cultured higher eukaryotic cells. Eukaryotic cells, particularly cultured cells of multicellular organisms, are preferred. Techniques for manipulating cloned DNA molecules and introducing exogenous DNA into a variety of host cells are disclosed by Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and

Ausubel et al. (eds.), Current Protocols in Molecular Biology, John Wiley and Sons, Inc., NY, 1987, which are incorporated herein by reference.

In general, a DNA sequence encoding a zFGF-10 polypeptide of the present invention is operably linked to other genetic elements required for its expression, generally including a transcription promoter and terminator within an expression vector. The vector will also commonly contain one or more selectable markers and one or more origins of replication, although those skilled in the art will recognize that within certain systems selectable markers may be provided on separate vectors, and replication of the exogenous DNA may be provided by integration into the host cell genome. Selection of promoters, terminators, selectable markers, vectors and other elements is a matter of routine design within the level of ordinary skill in the art. Many such elements are described in the literature and are available through commercial suppliers. Alternative markers that introduce an altered phenotype, such as green fluorescent protein, or cell surface proteins such as CD4, CD8, Class I MHC, placental alkaline phosphatase may be used to sort transfected cells from untransfected cells by such means as FACS sorting or magnetic bead separation technology.

To direct a zFGF-10 polypeptide into the secretory pathway of a host cell, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) is provided in the expression vector. The secretory signal sequence may be the native sequence, or a chimera comprising a signal sequence derived from another secreted protein (e.g., t-PA and α-pre-pro secretory leader) or synthesized *de novo*. The secretory signal sequence is joined to the zFGF-10 DNA sequence in the correct reading frame. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the polypeptide of interest, although certain signal sequences may be positioned elsewhere in the DNA sequence of interest (see, e.g., Welch et al., U.S. Patent No. 5,037,743; Holland et al., U.S. Patent No. 5,143,830).

Alternatively, the secretory signal sequence contained in the polypeptides of the present invention is used to direct other polypeptides into the secretory pathway. The present invention provides for such fusion polypeptides. A signal fusion polypeptide can be made wherein a secretory signal sequence derived from amino acid residues 1-17 of SEQ ID NO:2 is be operably linked to another polypeptide using methods known in the art and disclosed herein. The secretory signal sequence contained in the fusion polypeptides of the present invention is preferably fused amino-terminally to an additional peptide to direct the additional peptide into the secretory pathway. Such constructs have numerous applications known in the art. For example, these novel secretory signal sequence fusion constructs can direct the

secretion of an active component of a normally non-secreted protein. Such fusions may be used in vivo or in vitro to direct peptides through the secretory pathway.publication WO 94/06463. Insect cells can be infected with recombinant baculovirus, commonly derived from Autographa californica nuclear polyhedrosis virus (AcNPV). See, King, L.A. and Possee, R.D., The Baculovirus Expression System: A Laboratory Guide, London, Chapman & Hall; O'Reilly, D.R. et al., Baculovirus Expression Vectors: A Laboratory Manual, New York, Oxford University Press., 1994; and, Richardson, C. D., Ed., Baculovirus Expression Protocols. Methods in Molecular Biology, Totowa, NJ, Humana Press, 1995. A second method of making recombinant zFGF-10 baculovirus utilizes a transposon-based system described by Luckow (Luckow, V.A, et al., J Virol 67:4566-79, 1993). This system, which utilizes transfer vectors, is sold in the Bac-to-Bac™ kit (Life Technologies, Rockville, MD). This system utilizes a transfer vector, pFastBac1™ (Life Technologies) containing a Tn7 transposon to move the DNA encoding the zFGF-10 polypeptide into a baculovirus genome maintained in E. *coli* as a large plasmid called a "bacmid." See, Hill-Perkins, M.S. and Possee, R.D., J Gen Virol 71:971-6, 1990; Bonning, B.C. et al., J Gen Virol 75:1551-6, 1994; and, Chazenbalk, G.D., and Rapoport, B., J Biol Chem 270:1543-9, 1995. In addition, transfer vectors can include an in-frame fusion with DNA encoding an epitope tag at the C- or N-terminus of the expressed zFGF-10 polypeptide, for example, a Glu-Glu epitope tag (Grussenmeyer, T. et al., Proc. Natl. Acad. Sci. 82:7952-4, 1985). Using a technique known in the art, a transfer vector containing zFGF-10 is transformed into E. *coli*, and screened for bacmids which contain an interrupted lacZ gene indicative of recombinant baculovirus. The bacmid DNA containing the recombinant baculovirus genome is isolated, using common techniques, and used to transfect *Spodoptera frugiperda* cells, e.g. Sf9 cells. Recombinant virus that expresses zFGF-10 is subsequently produced. Recombinant viral stocks are made by methods commonly used the art.

The recombinant virus is used to infect host cells, typically a cell line derived from the fall armyworm, Spodoptera frugiperda. See, in general, Glick and Pasternak, Molecular Biotechnology: Principles and Applications of Recombinant DNA, ASM Press, Washington, D.C., 1994. Another suitable cell line is the High FiveO™ cell line (Invitrogen) derived from Trichoplusia ni (U.S. Patent No. 5,300,435). Commercially available serum-free media are used to grow and maintain the cells. Suitable media are Sf900 II™ (Life Technologies) or ESF 921™ (Expression Systems) for the Sf9 cells; and Ex-cell0405™ (JRH Biosciences, Lenexa, KS) or Express FiveO™ (Life Technologies) for the *T. ni* cells. The cells are grown up from an

inoculation density of approximately 2-5 x 10^5 cells to a density of 1-2 x 10^6 cells at which time a recombinant viral stock is added at a multiplicity of infection (MOI) of 0.1 to 10, more typically near 3. Procedures used are generally described in available laboratory manuals (King, L. A. and Possee, R.D., ibid.; O'Reilly, D.R. et al., ibid.; Richardson, C. D., ibid.). Subsequent purification of the zFGF-10 polypeptide from the supernatant can be achieved using methods described herein.

Fungal cells, including yeast cells, can also be used within the present invention. Yeast species of particular interest in this regard include *Saccharomyces cerevisiae, Pichia pastoris*, and *Pichia methanolica*. Methods for transforming S. *cerevisiae* cells with exogenous DNA and producing recombinant polypeptides therefrom are disclosed by, for example, Kawasaki, U.S. Patent No. 4,599,311; Kawasaki et al., U.S. Patent No. 4,931,373; Brake, U.S. Patent No. 4,870,008; Welch et al., U.S. Patent No. 5,037,743; and Murray et al., U.S. Patent No. 4,845,075. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient (e.g., leucine). A preferred vector system for use in *Saccharomyces cerevisiae* is the POT1 vector system disclosed by Kawasaki et al. (U.S. Patent No. 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. Suitable promoters and terminators for use in yeast include those from glycolytic enzyme 4,977,092) 132:3459-65, 1986 and Cregg, U.S. Patent No. 4,882,279. Aspergillus cells may be utilized 1986 and Cregg, U.S. Patent No. 4,882,279. Aspergillus cells may be utilized according to the methods of McKnight et al., U.S. Patent No. 4,935,349. Methods for transforming *Acremonium chrysogenum* are disclosed by Sumino et al., U.S. Patent No. 5,162,228. Methods for transforming *Neurospora* are disclosed by Lambowitz, U.S. Patent No. 4,486,533.

The use of *Pichia methanolica* as host for the production of recombinant proteins is disclosed in WIPO Publications WO 97/17450, WO 97/17451, WO 98/02536, and WO 98/02565. DNA molecules for use in transforming P. *methanolica* will commonly be prepared as double-stranded, circular plasmids, which are preferably linearized prior to transformation. For polypeptide production in P. *methanolica*, it is preferred that the promoter and terminator in the plasmid be that of a P. *methanolica* gene, such as a P. *methanolica* alcohol utilization gene (AUG1 or AUG2). Other useful promoters include those of the dihydroxyacetone synthase (DHAS), formate dehydrogenase (FMD), and catalase (CAT) genes. To facilitate integration of the DNA into the host chromosome, it is preferred to have the entire expression segment of the plasmid flanked at both ends by host DNA sequences. A preferred selectable marker

for use in *Pichia methanolica* is a P. *methanolica* ADE2 gene, which encodes phosphoribosyl-5-aminoimidazole carboxylase (AIRC; EC 4.1.1.21), which allows ade2 host cells to grow in the absence of adenine. For large-scale, industrial processes where it is desirable to minimize the use of methanol, it is preferred to use host cells in which both methanol utilization genes (AUG1 and AUG2) are deleted. For production of secreted proteins, host cells deficient in vacuolar protease genes (PEP4 and PRB1) are preferred. Electroporation is used to facilitate the introduction of a plasmid containing DNA encoding a polypeptide of interest into P. *methanolica* cells. It is preferred to transform P. *methanolica* cells by electroporation using an exponentially decaying, pulsed electric field having a field strength of from 2.5 to 4.5 kV/cm, preferably about 3.75 kV/cm, and a time constant (t) of from 1 to 40 milliseconds, most preferably about 20 milliseconds.

Cultured mammalian cells are suitable hosts within the present invention. Methods for introducing exogenous DNA into mammalian host cells include calcium phosphate-mediated transfection (Wigler et al., Cell 14:725, 1978; Corsaro and Pearson, Somatic Cell Genetics 7:603, 1981: Graham and Van der Eb, Virology 52:456, 1973), electroporation (Neumann et al., EMBO J. 1:841-5, 1982), DEAE-dextran mediated transfection (Ausubel et al., ibid.), and liposome-mediated transfection (Hawley-Nelson et al., Focus 15:73, 1993; Ciccarone et al., Focus 15:80, 1993, and viral vectors (Miller and Rosman, BioTechniques 7:980-90, 1989; Wang and Finer, Nature Med. 2:714-6, 1996). The production of recombinant polypeptides in cultured mammalian cells is disclosed, for example, by Levinson et al., U.S. Patent No. 4,713,339; Hagen et al., U.S. Patent No. 4,784,950; Palmiter et al., U.S. Patent No. 4,579,821; and Ringold, U.S. Patent No. 4,656,134. Suitable cultured mammalian cells include the COS-1 (ATCC No. CRL 1650), COS-7 (ATCC No. CRL 1651), BHK (ATCC No. CRL 1632), BHK 570 (ATCC No. CRL 10314), 293 (ATCC No. CRL 1573; Graham et al., J. Gen. Virol. 36:59-72, 1977) and Chinese hamster ovary (e.g. CHO-K1; ATCC No. CCL 61 or DG44) cell lines. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Rockville, Maryland. In general, strong transcription promoters are preferred, such as promoters from SV-40 or cytomegalovirus. See, e.g., U.S. Patent No. 4,956,288. Other suitable promoters include those from metallothionein genes (U.S. Patent Nos. 4,579,821 and 4,601,978)and the adenovirus major late promoter.

Drug selection is generally used to select for cultured mammalian cells into which foreign DNA has been inserted. Such cells are commonly referred to as "transfectants". Cells that have been cultured in the presence of the selective agent and

are able to pass the gene of interest to their progeny are referred to as "stable transfectants." A preferred selectable marker is a gene encoding resistance to the antibiotic neomycin. Selection is carried out in the presence of a neomycin-type drug, such as G-418 or the like. Selection systems can also be used to increase the expression level of the gene of interest, a process referred to as "amplification." Amplification is carried out by culturing transfectants in the presence of a low level of the selective agent and then increasing the amount of selective agent to select for cells that produce high levels of the products of the introduced genes. A preferred amplifiable selectable marker is dihydrofolate reductase, which confers resistance to methotrexate. Other drug resistance genes (e.g. hygromycin resistance, multi-drug resistance, puromycin acetyltransferase) can also be used.

Other higher eukaryotic cells can also be used as hosts, including plant cells, insect cells and avian cells. The use of Agrobacterium rhizogenes as a vector for expressing genes in plant cells has been reviewed by Sinkar et al., J. Biosci. (Bangalore) 11:47-58, 1987. Transformation of insect cells and production of foreign polypeptides therein is disclosed by Guarino et al., U.S. Patent No. 5,162,222 and WIPO

Transformed or transfected host cells are cultured according to conventional procedures in a culture medium containing nutrients and other components required for the growth of the chosen host cells. A variety of suitable media, including defined media and complex media, are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins and minerals. Media may also contain such components as growth factors or serum, as required. The growth medium will generally select for cells containing the exogenously added DNA by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker carried on the expression vector or co-transfected into the host cell. P. *methanolica* cells are cultured in a medium comprising adequate sources of carbon, nitrogen and trace nutrients at a temperature of about 25°C to 35°C. Liquid cultures are provided with sufficient aeration by conventional means, such as shaking of small flasks or sparging of fermentors. A preferred culture medium for P. *methanolica* is YEPD (2% D-glucose, 2% Bacto™ Peptone (Difco Laboratories, Detroit, MI), 1% Bacto™ yeast extract (Difco Laboratories), 0.004% adenine and 0.006% L-leucine).

It is preferred to purify the polypeptides of the present invention to ≥80% purity, more preferably to ≥90% purity, even more preferably ≥95% purity, and particularly preferred is a pharmaceutically pure state, that is greater than 99.9% pure with respect to contaminating macromolecules, particularly other proteins and nucleic

acids, and free of infectious and pyrogenic agents. Preferably, a purified polypeptide is substantially free of other polypeptides, particularly other polypeptides of animal origin.

Expressed recombinant zFGF-10 polypeptides (or chimeric zFGF-10 polypeptides) can be purified using fractionation and/or conventional purification methods and media. Ammonium sulfate precipitation and acid or chaotrope extraction may be used for fractionation of samples. Exemplary purification steps may include hydroxyapatite, size exclusion, FPLC and reverse-phase high performance liquid chromatography. Suitable anion exchange media include derivatized dextrans, agarose, cellulose, polyacrylamide, specialty silicas, and the like. PEI, DEAE, QAE and Q derivatives are preferred, with DEAE Fast-Flow Sepharose (Pharmacia, Piscataway, NJ) being particularly preferred. Exemplary chromatographic media include those media derivatized with phenyl, butyl, or octyl groups, such as Phenyl-Sepharose FF (Pharmacia), Toyopearl butyl 650 (Toso Haas, Montgomeryville, PA), Octyl-Sepharose (Pharmacia) and the like; or polyacrylic resins, such as Amberchrom CG 71 (Toso Haas) and the like. Suitable solid supports include glass beads, silica-based resins, cellulosic resins, agarose beads, cross-linked agarose beads, polystyrene beads, cross-linked polyacrylamide resins and the like that are insoluble under the conditions in which they are to be used. These supports may be modified with reactive groups that allow attachment of proteins by amino groups, carboxyl groups, sulfhydryl groups, hydroxyl groups and/or carbohydrate moieties. Examples of coupling chemistries include cyanogen bromide activation, N-hydroxysuccinimide activation, epoxide activation, sulfhydryl activation, hydrazide activation, and carboxyl and amino derivatives for carbodiimide coupling chemistries. These and other solid media are well known and widely used in the art, and are available from commercial suppliers. Methods for binding receptor polypeptides to support media are well known in the art. Selection of a particular method is a matter of routine design and is determined in part by the properties of the chosen support. See, for example, Affinity Chromatography: Principles & Methods, Pharmacia LKB Biotechnology, Uppsala, Sweden, 1988.

The polypeptides of the present invention can also be isolated by exploitation of their heparin binding properties. For a review, see, Burgess et al., Ann. Rev. of Biochem. 58:575-606, 1989. Members of the FGF family can be purified to apparent homogeneity by heparin-Sepharose affinity chromatography (Gospodarowicz et al., Proc. Natl. Acad. Sci. 81:6963-6967, 1984) and eluted using linear step gradients of NaCl (Ron et al., J. Biol. Chem. 268(4):2984-2988, 1993; Chromatography: Principles & Methods, pp. 77-80, Pharmacia LKB Biotechnology, Uppsala, Sweden, 1993; in "Immobilized Affinity Ligand Techniques", Hermanson et al., eds., pp. 165-

167, Academic Press, San Diego, 1992; Kjellen et al., Ann. Rev. Biochem.Ann. Rev. Biochem. 60:443-474, 1991; and Ke et al., Protein Expr. Purif. 3(6):497-507, 1992.)

Other purification methods include using immobilized metal ion adsorption (IMAC) chromatography to purify histidine-rich proteins. Briefly, a gel is first charged with divalent metal ions to form a chelate (E. Sulkowski, Trends in Biochem. 3:1-7, 1985). Histidine-rich proteins will be adsorbed to this matrix with differing affinities, depending upon the metal ion used, and will be eluted by competitive elution, lowering the pH, or use of strong chelating agents. Other methods of purification include purification of glycosylated proteins by lectin affinity chromatography and ion exchange chromatography (Methods in Enzymol., Vol. 182, "Guide to Protein Purification", M. Deutscher, (ed.), Acad. Press, San Diego, 1990, pp.529-39). Alternatively, a fusion of the polypeptide of interest and an affinity tag (e.g., polyhistidine, maltose-binding protein, an immunoglobulin domain) may be constructed to facilitate purification.

Protein refolding (and optionally reoxidation) procedures may be advantageously used. It is preferred to purify the protein to >80% purity, more preferably to >90% purity, even more preferably >95%, and particularly preferred is a pharmaceutically pure state, that is greater than 99.9% pure with respect to contaminating macromolecules, particularly other proteins and nucleic acids, and free of infectious and pyrogenic agents. Preferably, a purified protein is substantially free of other proteins, particularly other proteins of animal origin.

zFGF-10 polypeptides or fragments thereof may also be prepared through chemical synthesis. zFGF-10 polypeptides may be monomers or multimers; glycosylated or non-glycosylated; pegylated or non-pegylated; and may or may not include an initial methionine amino acid residue.

An in vivo approach for assaying proteins of the present invention involves viral delivery systems. Exemplary viruses for this purpose include adenovirus, herpesvirus, vaccinia virus and adeno-associated virus (AAV). Adenovirus, a double-stranded DNA virus, is currently the best studied gene transfer vector for delivery of heterologous nucleic acid (for a review, see T.C. Becker et al., Meth. Cell Biol. 43:161-89, 1994; and J.T. Douglas and D.T. Curiel, Science & Medicine 4:44-53, 1997). The adenovirus system offers several advantages: adenovirus can (i) accommodate relatively large DNA inserts; (ii) be grown to high-titer; (iii) infect a broad range of mammalian cell types; and (iv) be used with a large number of available vectors containing different promoters. Also, because adenoviruses are stable in the bloodstream, they can be administered by intravenous injection.

By deleting portions of the adenovirus genome, larger inserts (up to 7 kb) of heterologous DNA can be accommodated. These inserts can be incorporated into the viral DNA by direct ligation or by homologous recombination with a co-transfected plasmid. In an exemplary system, the essential E1 gene has been deleted from the viral vector, and the virus will not replicate unless the E1 gene is provided by the host cell (the human 293 cell line is exemplary). When intravenously administered to intact animals, adenovirus primarily targets the liver. If the adenoviral delivery system has an E1 gene deletion, the virus cannot replicate in the host cells. However, the host's tissue (e.g., liver) will express and process (and, if a secretory signal sequence is present, secrete) the heterologous protein. Secreted proteins will enter the circulation in the highly vascularized liver, and effects on the infected animal can be determined.

The adenovirus system can also be used for protein production in vitro. By culturing adenovirus-infected non-293 cells under conditions where the cells are not rapidly dividing, the cells can produce proteins for extended periods of time. For instance, BHK cells are grown to confluence in cell factories, then exposed to the adenoviral vector encoding the secreted protein of interest. The cells are then grown under serum-free conditions, which allows infected cells to survive for several weeks without significant cell division. Alternatively, adenovirus vector infected 293 cells can be grown as adherent cells or in suspension culture at relatively high cell density to produce significant amounts of protein (see Garnier et al., Cytotechnol. 15:145-55, 1994). With either protocol, an expressed, secreted heterologous protein can be repeatedly isolated from the cell culture supernatant. Within the infected 293S cell production protocol, non-secreted proteins may also be effectively obtained.

The activity of molecules of the present invention can be measured using a variety of assays that, for example, measure neogenesis or hyperplasia (i.e., proliferation) of neuronal, prostatic or pancreatic cells based on the tissue specificity. Additional activities likely associated with the polypeptides of the present invention include proliferation of endothelial cells, fibroblasts, skeletal myocytes, epithelial cells, epidermal cells and keratinocytes, directly or indirectly through other growth factors; action as a chemotaxic factor for endothelial cells, fibroblasts and/or phagocytic cells; osteogenic factor; and factor for expanding mesenchymal stem cell and precursor populations.

Proliferation can be measured using cultured cardiac cells or *in vivo* by administering molecules of the claimed invention to the appropriate animal model. Generally, proliferative effects are seen as an increase in cell number and therefore,

may include inhibition of apoptosis, as well as mitogenesis. Cultured cells include fibroblasts, skeletal myocytes, human umbilical vein endothelial cells from primary cultures. Established cell lines include: NIH 3T3 fibroblast (ATCC No. CRL-1658), CHH-1 chum heart cells (ATCC No. CRL-1680), H9c2 rat heart myoblasts (ATCC No. CRL-1446), Shionogi mammary carcinoma cells (Tanaka et al., Proc. Natl. Acad. Sci. 89:8928-8932, 1992) and LNCap.FGC adenocarcinoma cells (ATCC No. CRL-1740.) Assays measuring cell proliferation are well known in the art. For example, assays measuring proliferation include such assays as chemosensitivity to neutral red dye (Cavanaugh et al., Investigational New Drugs 8:347-354, 1990, incorporated herein by reference), incorporation of radiolabelled nucleotides (Cook et al., Analytical Biochem. 179:1- 7, 1989, incorporated herein by reference), incorporation of 5-bromo-2'-deoxyuridine (BrdU) in the DNA of proliferating cells (Porstmann et al., J. Immunol. Methods 82:169-179, 1985, incorporated herein by reference), and use of tetrazolium salts (Mosmann, J. Immunol. Methods 65:55-63, 1983; Alley et al., Cancer Res. 48:589-601, 1988; Marshall et al., Growth Reg. 5:69-84, 1995; and Scudiero et al., Cancer Res. 48:4827-4833, 1988; all incorporated herein by reference).

Differentiation is a progressive and dynamic process, beginning with pluripotent stem cells and ending with terminally differentiated cells. Pluripotent stem cells that can regenerate without commitment to a lineage express a set of differentiation markers that are lost when commitment to a cell lineage is made. Progenitor cells express a set of differentiation markers that may or may not continue to be expressed as the cells progress down the cell lineage pathway toward maturation. Differentiation markers that are expressed exclusively by mature cells are usually functional properties such as cell products, enzymes to produce cell products and receptors. The stage of a cell population's differentiation is monitored by identification of markers present in the cell population. Myocytes, osteoblasts, adipocytes, chrondrocytes, fibroblasts and reticular cells are believed to originate from a common mesenchymal stem cell (Owen et al., Ciba Fdn. Symp. 136:42-46, 1988). Markers for mesenchymal stem cells have not been well identified (Owen et al., J. of Cell Sci. 87:731-738, 1987), so identification is usually made at the progenitor and mature cell stages. The novel polypeptides of the present invention are useful for studies to isolate mesenchymal stem cells and myocyte progenitor cells, both *in vivo* and *ex vivo*.

There is evidence to suggest that factors that stimulate specific cell types down a pathway towards terminal differentiation or dedifferentiation, affects the entire cell population originating from a common precursor or stem cell. Thus, the present invention includes stimulation, inhibition, or proliferation of myocytes, smooth muscle

cells, osteoblasts, adipocytes, chondrocytes, neural tube-derived stem cells, neural crest stem cells, epidermal cells, keratinocytes, neuronal progenitors, pancreatic cells, prostate-derived cells and endothelial cells. Molecules of the present invention may, while stimulating proliferation or differentiation of cardiac myocytes, inhibit proliferation or differentiation of adipocytes, by virtue of the affect on their common precursor/stem cells. Thus molecules of the present invention, have use in inhibiting chondrosarcomas, atherosclerosis, restenosis and obesity.

Assays measuring differentiation include, for example, measuring cell-surface markers associated with stage-specific expression of a tissue, enzymatic activity, functional activity or morphological changes (Watt, FASEB, 5:281-284, 1991; Francis, Differentiation 57:63-75, 1994; Raes, Adv. Anim. Cell Biol. Technol. Bioprocesses, 161-171, 1989; all incorporated herein by reference).

In vivo assays for evaluating neogenesis or hyperplasia include cellular proliferation assays (Stern et al., Proc. Natl. Acad. Sci. 87:6808-6812, 1990, Lok et al., Nature 369:565-568, 1994), stimulation of the proliferation of neuronal and glial progenitors isolated from the septum and striatum (Palmer et al., Mol. Cell. Neurosci. 6:474-486, 1995), and stimulation of differentiation of neurons from neural crest progenitors (Vaisman et al., Development 115:1059-1069, 1992).

In vivo assays for measuring changes in bone formation rates include performing bone histology (see, Recker, R., eds. Bone Histomorphometry: Techniques and Interpretation. Boca Raton: CRC Press, Inc., 1983) and quantitative computed tomography (QCT; Ferretti,J. Bone 17:353S-364S, 1995; Orphanoludakis et al., Investig. Radiol. 14:122-130,, 1979 and Durand et al., Medical Physics 19:569-573, 1992). An *ex vivo* assay for measuring changes in bone formation would be, for example, a calavarial assay (Gowen et al., J. Immunol. 136:2478-2482, 1986).

With regard to modulating energy balance, particularly as it relates to adipocyte metabolism, proliferation and differentiation, zFGF-10 polypeptides may modulate effects on metabolic reactions. Such metabolic reactions include adipogenesis, gluconeogenesis, glycogenolysis, lipogenesis, glucose uptake, protein synthesis, thermogenesis, oxygen utilization and the like. Among other methods known in the art or described herein, mammalian energy balance may be evaluated by monitoring one or more of the aforementioned metabolic functions. These metabolic functions are monitored by techniques (assays or animal models) known to one of ordinary skill in the art, as is more fully set forth below. For example, the glucoregulatory effects of insulin are predominantly exerted in the liver, skeletal muscle

and adipose tissue. In skeletal muscle and adipose tissue, insulin acts to stimulate the uptake, storage and utilization of glucose.

Art-recognized methods exist for monitoring all of the metabolic functions recited above. Thus, one of ordinary skill in the art is able to evaluate zFGF-10 polypeptides, fragments, fusion proteins, antibodies, agonists and antagonists for metabolic modulating functions. Exemplary modulating techniques are set forth below.

Insulin-stimulated lipogenesis, for example, may be monitored by measuring the incorporation of ^{14}C-acetate into triglyceride (Mackall et al. J. Biol. Chem. 251:6462-6464, 1976) or triglyceride accumulation (Kletzien et al., Mol. Pharmacol. 41:393-398, 1992).

zFGF-10-stimulated uptake may be evaluated, for example, in an assay for insulin-stimulated glucose transport. Primary adipocytes or NIH 3T3 L1 cells (ATCC No. CCL-92.1) are placed in DMEM containing 1 g/l glucose, 0.5 or 1.0% BSA, 20 mM Hepes, and 2 mM glutamine. After two to five hours of culture, the medium is replaced with fresh, glucose-free DMEM containing 0.5 or 1.0% BSA, 20 mM Hepes, 1 mM pyruvate, and 2 mM glutamine. Appropriate concentrations of zFGF-10, insulin or IGF-1, or a dilution series of the test substance, are added, and the cells are incubated for 20-30 minutes. ^{3}H or ^{14}C-labeled deoxyglucose is added to ≈50 µM final concentration, and the cells are incubated for approximately 10-30 minutes. The cells are then quickly rinsed with cold buffer (e.g. PBS), then lysed with a suitable lysing agent (e.g. 1% SDS or 1 N NaOH). The cell lysate is then evaluated by counting in a scintillation counter. Cell-associated radioactivity is taken as a measure of glucose transport after subtracting non-specific binding as determined by incubating cells in the presence of cytochalasin b, an inhibitor of glucose transport. Other methods include those described by, for example, Manchester et al., Am. J. Physiol. 266 (Endocrinol. Metab. 29):E326-E333, 1994 (insulin-stimulated glucose transport).

Protein synthesis may be evaluated, for example, by comparing precipitation of ^{35}S-methionine-labeled proteins following incubation of the test cells with ^{35}S-methionine and ^{35}S-methionine and a putative modulator of protein synthesis.

Thermogenesis may be evaluated as described by B. Stanley in *The Biology of Neuropeptide Y and Related Peptides*, W. Colmers and C. Wahlestedt (eds.), Humana Press, Ottawa, 1993, pp. 457-509; C. Billington et al., Am. J. Physiol. 260:R321, 1991; N. Zarjevski et al., Endocrinology 133:1753, 1993; C. Billington et al., Am. J. Physiol. 266:R1765, 1994; Heller et al., Am. J. Physiol. 252(4 Pt 2): R661-7, 1987; and Heller et al., Am. J. Physiol. 245(3): R321-8, 1983. Also, metabolic rate,

which may be measured by a variety of techniques, is an indirect measurement of thermogenesis.

Oxygen utilization may be evaluated as described by Heller et al., Pflugers Arch. 369(1): 55-9, 1977. This method also involved an analysis of hypothalmic temperature and metabolic heat production. Oxygen utilization and thermoregulation have also been evaluated in humans as described by Haskell et al., J. Appl. Physiol. 51(4): 948-54, 1981.

zFGF-10 polypeptides can also be used to prepare antibodies that specifically bind to zFGF-10 epitopes, peptides or polypeptides. Methods for preparing polyclonal and monoclonal antibodies are well known in the art (see, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, NY, 1989; and Hurrell, J. G. R., Ed., Monoclonal Hybridoma Antibodies: Techniques and Applications, CRC Press, Inc., Boca Raton, FL, 1982, which are incorporated herein by reference). As would be evident to one of ordinary skill in the art, polyclonal antibodies can be generated from a variety of warm-blooded animals, such as horses, cows, goats, sheep, dogs, chickens, rabbits, mice, and rats.

The immunogenicity of a zFGF-10 polypeptide may be increased through the use of an adjuvant, such as alum (aluminum hydroxide) or Freund's complete or incomplete adjuvant. Polypeptides useful for immunization also include fusion polypeptides, such as fusions of zFGF-10 or a portion thereof with an immunoglobulin polypeptide or with maltose binding protein. The polypeptide immunogen may be a full-length molecule or a portion thereof. If the polypeptide portion is "hapten-like", such portion may be advantageously joined or linked to a macromolecular carrier (such as keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA) or tetanus toxoid) for immunization.

As used herein, the term "antibodies" includes polyclonal antibodies, affinity-purified polyclonal antibodies, monoclonal antibodies, and antigen-binding fragments, such as $F(ab')_2$ and Fab proteolytic fragments. Genetically engineered intact antibodies or fragments, such as chimeric antibodies, Fv fragments, single chain antibodies and the like, as well as synthetic antigen-binding peptides and polypeptides, are also included. Non-human antibodies may be humanized by grafting only non-human CDRs onto human framework and constant regions, or by incorporating the entire non-human variable domains (optionally "cloaking" them with a human-like surface by replacement of exposed residues, wherein the result is a "veneered" antibody). In some instances, humanized antibodies may retain non-human residues within the human variable region framework domains to enhance proper binding

characteristics. Through humanizing antibodies, biological half-life may be increased, and the potential for adverse immune reactions upon administration to humans is reduced. Alternative techniques for generating or selecting antibodies useful herein include *in vitro* exposure of lymphocytes to zFGF-10 protein or peptide, and selection of antibody display libraries in phage or similar vectors (for instance, through use of immobilized or labeled zFGF-10 protein or peptide).

Antibodies are defined to be specifically binding if they bind to a zFGF-10 polypeptide with a binding affinity (K_a) of 10^6 M^{-1} or greater, preferably 10^7 M^{-1} or greater, more preferably 10^8 M^{-1} or greater, and most preferably 10^9 M^{-1} or greater. The binding affinity of an antibody can be readily determined by one of ordinary skill in the art (for example, by Scatchard analysis).

A variety of assays known to those skilled in the art can be utilized to detect antibodies which specifically bind to zFGF-10 proteins or peptides. Exemplary assays are described in detail in Antibodies: A Laboratory Manual, Harlow and Lane (Eds.), Cold Spring Harbor Laboratory Press, 1988. Representative examples of such assays include: concurrent immunoelectrophoresis, radioimmunoassay, radioimmuno-precipitation, enzyme-linked immunosorbent assay (ELISA), dot blot or Western blot assay, inhibition or competition assay, and sandwich assay. In addition, antibodies can be screened for binding to wild-type versus mutant zFGF-10 protein or peptide.

Antibodies to zFGF-10 may be used for tagging cells that express zFGF-10; to target another protein, small molecule or chemical to heart tissue; for isolating zFGF-10 by affinity purification; for diagnostic assays for determining circulating levels of zFGF-10 polypeptides; for detecting or quantitating soluble zFGF-10 as marker of underlying pathology or disease; in analytical methods employing FACS; for screening expression libraries; for generating anti-idiotypic antibodies; and as neutralizing antibodies or as antagonists to block zFGF-10 mediated proliferation *in vitro* and *in vivo*. Suitable direct tags or labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent markers, chemiluminescent markers, magnetic particles and the like; indirect tags or labels may feature use of biotin-avidin or other complement/anti-complement pairs as intermediates. Antibodies herein may also be directly or indirectly conjugated to drugs, toxins, radionuclides and the like, and these conjugates used for *in vivo* diagnostic or therapeutic applications.

Molecules of the present invention can be used to identify and isolate receptors involved in neuronal or pancreatic cell proliferation. For example, proteins and peptides of the present invention can be immobilized on a column and membrane preparations run over the column (Immobilized Affinity Ligand Techniques,

Hermanson et al., eds., Academic Press, San Diego, CA, 1992, pp.195-202). Proteins and peptides can also be radiolabeled (Methods in Enzymol., vol. 182, "Guide to Protein Purification", M. Deutscher, ed., Acad. Press, San Diego, 1990, 721-737) or photoaffinity labeled (Brunner et al., Ann. Rev. Biochem. 62:483-514, 1993 and Fedan et al., Biochem. Pharmacol. 33:1167-1180, 1984) and specific cell-surface proteins can be identified.

Antagonists will be useful for inhibiting the proliferative activities of zFGF-10 molecules, in cell types such as neuronal, pancreatic, epithelial cells, epidermal cells, keratinocytes, and prostatic cells, including fibroblasts and endothelial cells. For example, antagonists to zFGF-10 will be useful for inhibitions of disorders associated with kidney epithelium, such as glomerulonephritis. Disorders associated with keratinocytes, such as psoriasis may be inhibited by zFGF-10 antagonists. Genes encoding zFGF-10 polypeptide binding domains can be obtained by screening random peptide libraries displayed on phage (phage display) or on bacteria, such as *E. coli.* Nucleotide sequences encoding the polypeptides can be obtained in a number of ways, such as through random mutagenesis and random polynucleotide synthesis. These random peptide display libraries can be used to screen for peptides which interact with a known target which can be a protein or polypeptide, such as a ligand or receptor, a biological or synthetic macromolecule, or organic or inorganic substances. Techniques for creating and screening such random peptide display libraries are known in the art (Ladner et al., US Patent NO:5,223,409; Ladner et al., US Patent NO:4,946,778; Ladner et al., US Patent NO:5,403,484 and Ladner et al., US Patent NO:5,571,698) and random peptide display libraries and kits for screening such libraries are available commercially, for instance from Clontech (Palo Alto, CA), Invitrogen Inc. (San Diego, CA), New England Biolabs, Inc. (Beverly, MA) and Pharmacia LKB Biotechnology Inc. (Piscataway, NJ). Random peptide display libraries can be screened using the zFGF-10 sequences disclosed herein to identify proteins which bind to zFGF-10. These "binding proteins" which interact with zFGF-10 polypeptides may be used for tagging cells; for isolating homolog polypeptides by affinity purification; they can be directly or indirectly conjugated to drugs, toxins, radionuclides and the like. These binding proteins can also be used in analytical methods such as for screening expression libraries and neutralizing activity. The binding proteins can also be used for diagnostic assays for determining circulating levels of polypeptides; for detecting or quantitating soluble polypeptides as marker of underlying pathology or disease. These binding proteins can also act as zFGF "antagonists" to block zFGF-10 binding and signal transduction *in vitro* and *in vivo*. These anti- zFGF-10 binding proteins would be useful

for inhibiting expression of genes which result in proliferation or differentiation. Such anti-zFGF-10 binding proteins can be used for treatment, for example, in neuroblastoma, glioblastoma, prostatic hypertrophy, prostatic carcinoma, pancreatic carcinoma, and spinal cord injury, alone or combination with other therapies.

The molecules of the present invention will be useful for proliferation of neuronal, prostatic and pancreatic tissue cells, such as pancreatic islets, pancreatic acinar cells, neuroectoderm, neurons of the central nervous systems, and sympathetic neurons *in vitro*. For example, molecules of the present invention are useful as components of defined cell culture media, and may be used alone or in combination with other cytokines and hormones to replace serum that is commonly used in cell culture. Molecules of the present invention are particularly useful in specifically promoting the growth and/or development of pancreatic islets, prostate cells (e.g., PZ-HPV-7 human prostate epithelium cells ATCC Number: CRL-2221 and rat YPEN-1 normal prostate cells, ATCC Number: CRL-2222); neuronal cells (e.g., mouse CATH.a brain neuronal cells ATCC Number: CRL- 11179, human HCN-1A neuronal cells, ATCC Number: CRL-10442) in culture, and may also prove useful in the study of hyperplasia and regeneration. Other types of cells for which zFGF-10 molecules will be useful for establishing and maintaining cell cultures include epithelial cells and keratinocytes. Epithelial cells can be isolated from, for example, prostate, cornea, lung, mammary or kidney tissues.

The polypeptides, nucleic acid and/or antibodies of the present invention may be used in treatment of disorders associated with diabetes mellitus, neural cell development or degeneration, amyotrophic lateral sclerosis, cerebrovascular stroke, neurophathy associated with lack of maintenance of neuronal differentiation, and congenital disorders of the nervous system or lack of neuronal development. Molecules of the present invention may also be useful for promoting angiogenesis and wound healing, for revascularization in the eye, for complications related to poor circulation such as diabetic foot ulcers, for stroke, following coronary reperfusion using pharmacologic methods and other indications where angiogenesis is of benefit. Molecules of the present invention may be useful for improving cardiac function, either by inducing cardiac myocyte neogenesis and/or hyperplasia, by inducing coronary collateral formation, or by inducing remodeling of necrotic myocardial area.

ZFGF-10 will be useful for promoting wound healing of the epidermis. The molecules of the present invention can be used to protect and promote recovery of the epithelial cells in the gastrointestinal tract, small intestine and oral muscosa after treat with chemotherapy and/or radiation. Stimulation of lung epithelial cells lining the

air space can promote recovery from lung injury and complications associated with premature birth in neonates. ZFGF-10 may also modulate surfactant production in the lung epithelium. Other epithelial cells are found in prostate, cornea, mammary and kidney tissue, and the proliferation and specialized cell functions of these cells can be modulated by zFGF-10.

An ischemic event is the disruption of blood flow to an organ, resulting in necrosis or infarct of the non-perfused region. Ischemia-reperfusion is the interruption of blood flow to an organ, such as the heart or brain, and subsequent restoration (often abrupt) of blood flow. While restoration of blood flow is essential to preserve functional tissue, the reperfusion itself is known to be deleterious. In fact, there is evidence that reperfusion of an ischemic area compromises endothelium-dependent vessel relaxation resulting in vasospasms, and in the heart compromised coronary vasodilation, that is not seen in an ischemic event without reperfusion (Cuevas et al., Growth Factors 15:29-40, 1997). Both ischemia and reperfusion are important contributors to tissue necrosis, such as a myocardial infarct or stroke. The molecules of the present invention will have therapeutic value to reduce damage to the tissues caused by ischemia or ischemia-reperfusion events, particularly in the heart or brain.

Other therapeutic uses for the present invention include induction of skeletal muscle neogenesis and/or hyperplasia, kidney regeneration and/or for treatment of systemic and pulmonary hypertension.

ZFGF-10 induced coronary collateral development is measured in rabbits, dogs or pigs using models of chronic coronary occlusion (Landau et al., Amer. Heart J. 29:924-931, 1995; Sellke et al., Surgery 120(2):182-188, 1996 and Lazarous et al., 1996, ibid.) zFGF-10 benefits for treating stroke is tested *in vivo* in rats utilizing bilateral carotid artery occlusion and measuring histological changes, as well as maze performance (Gage et al., Neurobiol. Aging 9:645-655, 1988). ZFGF-10 efficacy in hypertension is tested *in vivo* utilizing spontaneously hypertensive rats (SHR) for systemic hypertension (Marche et al., Clin. Exp. Pharmacol. Physiol. Suppl. 1:S114-116, 1995).

Molecules of the present invention can be used to target the delivery of agents or drugs to the cells and/or tissues derived from the neuroectoderm, the developing central nervous systems, the developing peripheral nervous system, the developing spinal cord, prostate and pancreas. For example, the molecules of the present invention will be useful limiting expression to the neural tissue, by virtue of the tissue specific expression directed by the zFGF-10 promoter. For example, neural tissue-specific expression can be achieved using a zFGF-10-adenoviral discistronic

construct (Rothmann et al., Gene Therapy 3:919-926, 1996). In addition, the zFGF-10 polypeptides can be used to restrict other agents or drugs to neural tissue by linking zFGF-10 polypeptides to another protein (Franz et al., Circ. Res. 73:629-638, 1993) by linking a first molecule that is comprised of a zFGF-10 homolog polypeptide with a second agent or drug to form a chimera. Proteins, for instance antibodies, can be used to form chimeras with zFGF-10 molecules of the present invention (Narula et al., J. Nucl. Cardiol. 2:26-34, 1995). Examples of agents or drugs include, but are not limited to, bioactive-polypeptides, genes, toxins, radionuclides, small molecule pharmaceuticals and the like. Linking may be direct or indirect (e.g., liposomes), and may occur by recombinant means, chemical linkage, strong non-covalent interaction and the like.

For pharmaceutical use, the proteins of the present invention are formulated for parenteral, particularly intravenous or subcutaneous, administration according to conventional methods. Intravenous administration will be by bolus injection or infusion over a typical period of one to several hours. In general, pharmaceutical formulations will include a zFGF-10 protein in combination with a pharmaceutically acceptable vehicle, such as saline, buffered saline, 5% dextrose in water or the like. Formulations may further include one or more excipients, preservatives, solubilizers, buffering agents, albumin to prevent protein loss on vial surfaces, etc. Methods of formulation are well known in the art and are disclosed, for example, in Remington's Pharmaceutical Sciences, Gennaro, ed., Mack Publishing Co., Easton PA, 1990, which is incorporated herein by reference. Therapeutic doses will generally be in the range of 0.1 to 100 μg/kg of patient weight per day, preferably 0.5-20 μg/kg per day, with the exact dose determined by the clinician according to accepted standards, taking into account the nature and severity of the condition to be treated, patient traits, etc. Determination of dose is within the level of ordinary skill in the art. The proteins may be administered for acute treatment, over one week or less, often over a period of one to three days or may be used in chronic treatment, over several months or years. In general, a therapeutically effective amount of zFGF-10 is an amount sufficient to produce a clinically significant change in proliferation, or increases in specific cell types associated with mesenchymal stem cells and progenitors.

. ZFGF10 polypeptides can also be used to teach analytical skills such as mass spectrometry, circular dichroism, to determine conformation, especially of the four alpha helices, x-ray crystallography to determine the three-dimensional structure in atomic detail, nuclear magnetic resonance spectroscopy to reveal the structure of proteins in solution. For example, a kit containing the ZFGF10 can be given to the

student to analyze. Since the amino acid sequence would be known by the instructor, the protein can be given to the student as a test to determine the skills or develop the skills of the student, the instructor would then know whether or not the student has correctly analyzed the polypeptide. Since every polypeptide is unique, the educational utility of ZFGF10 would be unique unto itself.

The antibodies which bind specifically to ZFGF10 can be used as a teaching aid to instruct students how to prepare affinity chromatography columns to purify ZFGF10, cloning and sequencing the polynucleotide that encodes an antibody and thus as a practicum for teaching a student how to design humanized antibodies. The *ZFGF10* gene, polypeptide, or antibody would then be packaged by reagent companies and sold to educational institutions so that the students gain skill in art of molecular biology. Because each gene and protein is unique, each gene and protein creates unique challenges and learning experiences for students in a lab practicum. Such educational kits containing the *ZFGF10* gene, polypeptide, or antibody are considered within the scope of the present invention. The invention is further illustrated by the following non-limiting examples.

In summary, the present invention provides isolated polypeptides comprising a sequence of amino acid residues that is at least 95% identical to the sequence as shown in SEQ ID NO: 2 from amino acid residue 18 to amino acid residue 170. In other embodiments, the polypeptides include a secretory signal sequence, such as the amino acid sequence shown in SEQ ID NO: 2 from residue 1 to residue 17. In other embodiments, the polypeptide will be at least 95% identical to SEQ ID NO: 2 from residue 18 to residue 170, and will include a cysteine at position 113, a phenylalanine at position 115, and glutamic acid at position 117. In other embodiments, the polypeptide will also include a leucine at position 53, a valine at position 61, a leucine at position 74, an isoleucine at position 76, a valine at position 84, an isoleucine at position 86, a cysteine at position 113, a phenylalanine at position 115, a glutamic acid at position 117, an isoleucine at position 119, a tyrosine at position 127, and leucine at position 141, a phenylalanine at position 164 as shown in SEQ ID NO: 2. In another embodiment, the polypeptides will be at 15 continuous amino acids of the sequence as shown in SEQ ID NO: 2.

Another aspect of the present invention provides expression vectors comprising operably linked elements that include a transcription promoter; and DNA segment as described above, and a transcription terminator.

In another aspect, the present invention includes cultured cells expressing the expression vectors described herein. Other aspects include methods of

making the polypeptides described herein by expressing the polynucleotide encoding for the polypeptides described herein in cultured cells, and then recovering the polypeptides. Antibodies to zFGF10 polypeptides as described herein are provided as well.

Additional aspects of the present invention include isolated polynucleotides comprising a sequence of nucleotides that encode for a polypeptide that is at least 95% identical to the sequence of amino acid residues as shown in SEQ ID NO: 2 from amino acid residue 180 through amino acid residue 170 and described herein. In other embodiments, the present invention includes isolated polynucleotide molecules that comprises a sequence of nucleotides as shown in either SEQ ID NO: 1 or SEQ ID NO: 3 from nucleotide 52 to nucleotide 510. In other embodiments, the isolated polynucleotide molecules comprises a sequence of nucleotides as shown in either SEQ ID NO: 1 or SEQ ID NO: 3 from nucleotide 1 to nucleotide 510.

Furthermore, the present invention provides methods for stimulating the proliferation of various mesenchymal cells and mesenchymal stem cells and progenitors that comprise culturing these cells in the presence of zFGF10 polypeptides as described herein in an amount sufficient to increase the number of cells when compared to cells grown in the absence of zFGF10 polypeptide. Other methods include stimulating the proliferation of epidermal cells and keratinocytes and their progenitors.

EXAMPLES

Example 1

Homologous recombination in yeast is used to create expression plasmids containing the polynucleotide encoding zFGF-10 for expression in mammalian cells. To construct the zFGF-10/pCZF199 expression vectors the following DNA fragments are transformed into *S. cerevisiae*: Sna BI digested pCZR199 as an acceptor vector, the zFGF-10 EcoRI restriction fragment, and two, double stranded linker segments. The expression vector, pCZR199, has yeast replication elements, (CEN, ARS), the selectable marker, URA3, *E. coli* replication elements (e.g., AMP^R and *ori*), a blunt-ended cloning site, Sna BI, and adds either a N-terminal or C-terminal Glu-Glu tag (SEQ ID NO: 15). The vectors are used to create zFGF-10 polypeptides having either end of the expressed protein Glu-Glu tagged. The double stranded linker segments are prepared using PCR. The linkers served to join the vector to the insert fragments at both the 5' and 3' ends. Two sets of linkers are prepared. One set of linkers joins the insert to a vector placing the Glu-Glu tag (SEQ ID NO: 15)

on the 5' end of the insert sequence using a linker. The second set of linkers is used to join the zFGF-10 insert into a vector placing a 3' Glu-Glu tag (SEQ ID NO: 15). A third set of linkers is used to join the zFGF-10 insert into the vector, resulting in an untagged constructs The 5' linker is same as the linked used for the C-terminally Glu-Glu tagged zFGF-10. The 3' linker is the same as the linker used for the N-terminally Glu-Glu tagged zFGF-10. The oligonucleotides are joined using standard PCR reaction conditions and heated to 94°C for 1.5 minutes followed by 10 cycles at 94°C for 30 seconds, 50°C for 1 minute and 72°C for 1 minute, then a 10 minute extension at 72°C.

The DNA fragments are added to 100 µl competent yeast (Genetic strain SF838-9Dα, Roffman et al., EMBO J. 8:2057-65, 1989) and electroporated. The yeast cells are immediately diluted in 600 µl 1.2 M sorbitol and plated on Ura D plates and incubated at 30oC for 48 hours. Ura^+ colonies are selected from both the N-terminally-tagged and C-terminally-tagged zFGF-10 proteins and the DNA from the resulting yeast colonies is extracted and transformed into *E. coli*. Individual clones harboring the correct expression construct are identified by restriction digests. DNA sequencing confirms that the desired sequences has been enjoined with one another.

Large scale plasmid DNA is isolated from one or more correct clones from both the N- and C-terminally tagged zFGF-10 sequences, the expression cassette liberated from the vector and transformed into yeast or *E. coli* for large scale protein production.

Example 2

The procedure described below is used for protein expressed in conditioned media of E. *coli*, *Pichia methanolica*, and chinese hamster ovary cells (CHO). For zFGF-10 expressed in *E. coli* and *Pichia*, however, the media is not concentrated before application to the AF Heparin 650m affinity column. Unless otherwise noted, all operations are carried out at 4°C. A total of 25 liters of conditioned media from CHO cells is sequentially sterile filtered through a 4 inch, 0.2 mM Millipore (Bedford, MA) OptiCap capsule filter and a 0.2 mM Gelman (Ann Arbor, MI) Supercap 50. The material is then concentrated to about 1.3 liters using a Millipore ProFlux A30 tangential flow concentrator fitted with a 3000 kDa cutoff Amicon (Bedford, MA) S10Y3 membrane. The concentrated material is again sterile-filtered with the Gelman filter as described above. A mixture of protease inhibitors is added to the concentrated conditioned media to final concentrations of 2.5 mM ethylenediaminetetraacetic acid (EDTA, Sigma Chemical Co. St. Louis, MO), 0.001

mM leupeptin (Boehringer-Mannheim, Indianapolis, IN), 0.001 mM pepstatin (Boehringer-Mannheim) and 0.4 mM Pefabloc (Boehringer-Mannheim).

The concentrated conditioned media is applied to a 5.0 x 15.0 cm AF Heparin 650m (TosoHaas, Montgomeryville, PA) column equilibrated in 0.25M NaCl, 50 mM sodium phosphate, pH 7.2 at a flow rate of 5 ml/min using a BioCad Sprint HPLC (PerSeptive BioSystems, Framingham, MA). Two-ml fractions are collected and the absorbance at 280 nM is monitored. After sample application, the column is washed with 10 column volumes of loading buffer and when the absorbance of the effluent is less than that 0.05, the column is eluted with a three column volume gradient from 0.25 M to 2.0 M NaCl in 50 mM sodium phosphate, pH 7.2. The fractions containing zFGF-10 are identified by SDS-PAGE and western blotting with anti-zFGF-10 antibodies.

Fractions containing zFGF-10 are pooled together and diluted ten-fold into 50 mM sodium phosphate pH 7.5 and the material is applied to a 1.5 x 20.0 cm Poros HS cation exchange column equilibrated in 50 mM phosphate pH 7.5 using the BioCad Sprint as described above. After sample application, the column is washed with 10 column volumes of loading buffer and when the absorbance of the effluent is less than that 0.05, the column is eluted with a 40.0 column volume gradient from 0.0 M to 2.0 M NaCl in 50 mM sodium phosphate, pH 7.5. Fractions are collected as described above and those containing zFGF-10 will be identified by SDS-PAGE and Western blotting, pooled together and concentrated using an Amicon stirred cell fitted with a YM-10 membrane.

The concentrated material is then be applied to a 3.5 x 100 cm Sephacryl-S100 gel filtration column equilibrated in 1.0 M NaCl, 0.01 M EDTA and 0.05 M sodium phosphate, pH 7.2. Fractions are analyzed by SDS-PAGE and Western blotting with anti-zFGF-10 antibodies as described above. Fractions containing pure zFGF-10 are pooled together and samples are taken for amino acid analysis and N-terminal sequencing. The remainder of the sample is aliquoted, and stored at -80°C.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

We claim:

1. An isolated polypeptide comprising a sequence of amino acid residues that is at least 95% identical to the sequence as shown in SEQ ID NO:2 from residue 18 through residue 170.

2. The isolated polypeptide of claim 1, wherein the polypeptide comprises a Cys residue at position 113, a Phe residue at position 115, and a Glu residue at position 117, as shown in SEQ ID NO:2.

3. The isolated polypeptide of claim 1 wherein the polypeptide comprises a Leu residue at position 53, a Val residue at position 61, a Leu residue at position 74, a Ile residue at position 76, a Val residue at position 84, a Ile residue at position 86, a Val residue at position 95, a Leu residue at position 103, Cys residue at position 113, a Phe residue at position 115, Glu residue at position 117, a Ile residue at position 119, a Tyr residue at position 127, a Leu residue at position 141, and a Phe residue at position 164, as shown in SEQ ID NO:2.

4. An isolated polypeptide comprising a sequence of amino acid residues as shown in SEQ ID NO:2 from residue 18 through residue 170.

5. An isolated polypeptide comprising at least 15 contiguous amino acid residues of SEQ ID NO:2.

6. An expression vector comprising the following operably linked elements:

(a) a transcription promoter;

(b) a DNA segment encoding a polypeptide according to claim 1; and

(c) a transcription terminator.

7. The expression vector of claim 6 wherein a secretory signal sequence is operably linked to the DNA segment.

8. An expression vector comprising the following operably linked elements:

(a) a transcription promoter;

(b) a DNA segment encoding a polypeptide according to claim 4; and

(c) a transcription terminator.

9. The expression vector of claim 8, wherein a secretory signal sequence is operably linked to the DNA segment.

10. The expression vector of claim 9, wherein the secretory signal sequence is polynucleotide that encodes for the amino acid sequence as shown in SEQ ID NO: 2 from amino acid residue 1 (Met) to amino acid residue 17 (Ala).

11. A cultured cell comprising the expression vector of claim 6.

12. A cultured cell comprising the expression vector of claim 8.

13. A method of making a polypeptide comprising:

culturing a cell according to claim 12 under conditions wherein the DNA segment is expressed; and

recovering the protein encoded by the DNA segment.

14. An antibody that specifically binds to the polypeptide of claim 4 or a protein comprising the polypeptide of claim 4.

15. An isolated polynucleotide molecule comprising a sequence of nucleotides that encodes for a sequence of amino acid residues that is at least 95% identical to the sequence of amino acid residues as shown in SEQ ID NO:2 from amino acid residue 18 through amino acid residue 170.

16. An isolated polynucleotide molecule comprising a sequence of nucleotides as shown in SEQ ID NO: 1 from nucleotide 52 to nucleotide 510 or SEQ ID NO: 3 from nucleotide 52 to nucleotide 510.

17. The isolated polynucleotide molecule of claim 16, wherein the sequence of nucleotides is from nucleotide 1 to nucleotide 510 of SEQ ID NO: 1, or nucleotide 1 to nucleotide 510 of SEQ ID NO: 3.

18. A method for stimulating proliferation of mesenchymal cells comprising culturing mesenchymal stem cells or progenitor cells in the presence of zFGF10 polypeptide comprising a sequence of amino acid residues as shown in SEQ ID NO:2 from residue 18 through residue 170, in an amount sufficient to increase the number of mesenchymal cells as compared to cells grown the absence of zFGF10 polypeptide.

19. A method of stimulating the proliferation of epidermal cells comprising culturing epidermal cells or epidermal cell progenitors in the presence of zFGF10 polypeptide comprising a sequence of amino acid residues as shown in SEQ ID NO:2 from residue 18 through residue 170, in an amount sufficient to increase the number of epidermal cells as compared to cells grown in the absence of zFGF10 polypeptide.

20. A method of stimulating the proliferation of keratinocytes comprising culturing keratinocytes or keratinocyte progenitors in the presence of zFGF10 polypeptide comprising a sequence of amino acid residues as shown in SEQ ID NO:2 from residue 18 through residue 170, in an amount sufficient to increase the number of keratinocytes as compared to cells grown the in absence of zFGF10 polypeptide.

SEQUENCE LISTING

<110> ZymoGenetics, Inc.

<120> NOVEL FGF HOMOLOG ZFGF10

<130> 99-83PC

<150> US 09/474,279
<151> 1999-12-29

<160> 15

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 513
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)...(513)

<400> 1

```
atg cgc cgc cgc ctg tgg ctg ggc ctg gcc tgg ctg ctg ctg gcg cgg       48
Met Arg Arg Arg Leu Trp Leu Gly Leu Ala Trp Leu Leu Leu Ala Arg
 1               5                  10                  15

gcg ccg gac gcc gcg gga acc ccg agc gcg tcg cgg gga ccg cgc agc       96
Ala Pro Asp Ala Ala Gly Thr Pro Ser Ala Ser Arg Gly Pro Arg Ser
            20                  25                  30

tac ccg cac ctg gag ggc gac gtg cgc tgg cgg cgc ctc ttc tcc tcc      144
Tyr Pro His Leu Glu Gly Asp Val Arg Trp Arg Arg Leu Phe Ser Ser
        35                  40                  45

act cac ttc ttc ctg cgc gtg gat ccc ggc ggc cgc gtg cag ggc acc      192
Thr His Phe Phe Leu Arg Val Asp Pro Gly Gly Arg Val Gln Gly Thr
    50                  55                  60
```

```
cgc tgg cgc cac ggc cag gac agc atc ctg gag atc cgc tct gta cac      240
Arg Trp Arg His Gly Gln Asp Ser Ile Leu Glu Ile Arg Ser Val His
 65                  70                  75                  80

gtg ggc gtc gtg gtc atc aaa gca gtg tcc tca ggc ttc tac gtg gcc      288
Val Gly Val Val Val Ile Lys Ala Val Ser Ser Gly Phe Tyr Val Ala
                 85                  90                  95

atg aac cgc cgg ggc cgc ctc tac ggg tcg cga ctc tac acc gtg gac      336
Met Asn Arg Arg Gly Arg Leu Tyr Gly Ser Arg Leu Tyr Thr Val Asp
            100                 105                 110

tgc agg ttc cgg gag cgc atc gaa gag aac ggc cac aac acc tac gcc      384
Cys Arg Phe Arg Glu Arg Ile Glu Glu Asn Gly His Asn Thr Tyr Ala
        115                 120                 125

tca cag cgc tgg cgc cgc cgc ggc cag ccc atg ttc ctg gcg ctg gac      432
Ser Gln Arg Trp Arg Arg Arg Gly Gln Pro Met Phe Leu Ala Leu Asp
    130                 135                 140

agg agg ggg ggg ccc cgg cca ggc ggc cgg acg cgg cgg tac cac ctg      480
Arg Arg Gly Gly Pro Arg Pro Gly Gly Arg Thr Arg Arg Tyr His Leu
145                 150                 155                 160

tcc gcc cac ttc ctg ccc gtc ctg gtc tcc tga                          513
Ser Ala His Phe Leu Pro Val Leu Val Ser  *
                165                 170
```

<210> 2
<211> 170
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Arg Arg Arg Leu Trp Leu Gly Leu Ala Trp Leu Leu Leu Ala Arg
 1               5                  10                  15
Ala Pro Asp Ala Ala Gly Thr Pro Ser Ala Ser Arg Gly Pro Arg Ser
            20                  25                  30
Tyr Pro His Leu Glu Gly Asp Val Arg Trp Arg Arg Leu Phe Ser Ser
        35                  40                  45
Thr His Phe Phe Leu Arg Val Asp Pro Gly Gly Arg Val Gln Gly Thr
    50                  55                  60
```

```
Arg Trp Arg His Gly Gln Asp Ser Ile Leu Glu Ile Arg Ser Val His
65                  70                  75                  80
Val Gly Val Val Val Ile Lys Ala Val Ser Ser Gly Phe Tyr Val Ala
                85                  90                  95
Met Asn Arg Arg Gly Arg Leu Tyr Gly Ser Arg Leu Tyr Thr Val Asp
            100                 105                 110
Cys Arg Phe Arg Glu Arg Ile Glu Glu Asn Gly His Asn Thr Tyr Ala
        115                 120                 125
Ser Gln Arg Trp Arg Arg Arg Gly Gln Pro Met Phe Leu Ala Leu Asp
    130                 135                 140
Arg Arg Gly Gly Pro Arg Pro Gly Gly Arg Thr Arg Arg Tyr His Leu
145                 150                 155                 160
Ser Ala His Phe Leu Pro Val Leu Val Ser
                165                 170
```

<210> 3
<211> 510
<212> DNA
<213> Artificial Sequence

<220>
<223> degenerate sequence

<221> misc_feature
<222> (1)...(510)
<223> n = A,T,C or G

<400> 3

```
atgmgnmgnm gnytntggyt nggnytngcn tggytnytny tngcnmgngc nccngaygcn     60
gcnggnacnc cnwsngcnws nmgnggnccn mgnwsntayc cncayytnga rggngaygtn    120
mgntggmgnm gnytnttyws nwsnacncay ttyttyytnm gngtngaycc nggnggnmgn    180
gtncarggna cnmgntggmg ncayggncar gaywsnathy tngarathmg nwsngtncay    240
gtnggngtng tngtnathaa rgcngtnwsn wsnggnttyt aygtngcnat gaaymgnmgn    300
ggnmgnytnt ayggnwsnmg nytntayacn gtngaytgym gnttymgnga rmgnathgar    360
garaayggnc ayaayacnta ygcnwsncar mgntggmgnm gnmgnggnca rccnatgtty    420
ytngcnytng aymgnmgngg nggnccnmgn ccnggnggnm gnacnmgnmg ntaycayytn    480
wsngcncayt tyytnccngt nytngtnwsn                                     510
```

<210> 4
<211> 208
<212> PRT
<213> Homo sapiens

<400> 4
Met Trp Lys Trp Ile Leu Thr His Cys Ala Ser Ala Phe Pro His Leu
1 5 10 15
Pro Gly Cys Cys Cys Cys Cys Phe Leu Leu Leu Phe Leu Val Ser Ser
20 25 30
Val Pro Val Thr Cys Gln Ala Leu Gly Gln Asp Met Val Ser Pro Glu
35 40 45
Ala Thr Asn Ser Ser Ser Ser Ser Phe Ser Ser Pro Ser Ser Ala Gly
50 55 60
Arg His Val Arg Ser Tyr Asn His Leu Gln Gly Asp Val Arg Trp Arg
65 70 75 80
Lys Leu Phe Ser Phe Thr Lys Tyr Phe Leu Lys Ile Glu Lys Asn Gly
85 90 95
Lys Val Ser Gly Thr Lys Lys Glu Asn Cys Pro Tyr Ser Ile Leu Glu
100 105 110
Ile Thr Ser Val Glu Ile Gly Val Val Ala Val Lys Ala Ile Asn Ser
115 120 125
Asn Tyr Tyr Leu Ala Met Asn Lys Lys Gly Lys Leu Tyr Gly Ser Lys
130 135 140
Glu Phe Asn Asn Asp Cys Lys Leu Lys Glu Arg Ile Glu Glu Asn Gly
145 150 155 160
Tyr Asn Thr Tyr Ala Ser Phe Asn Trp Gln His Asn Gly Arg Gln Met
165 170 175
Tyr Val Ala Leu Asn Gly Lys Gly Ala Pro Arg Arg Gly Gln Lys Thr
180 185 190
Arg Arg Lys Asn Thr Ser Ala His Phe Leu Pro Met Val Val His Ser
195 200 205

<210> 5
<211> 194
<212> PRT
<213> Homo sapins

<400> 5
Met His Lys Trp Ile Leu Thr Trp Ile Leu Pro Thr Leu Leu Tyr Arg
1 5 10 15
Ser Cys Phe His Ile Ile Cys Leu Val Gly Thr Ile Ser Leu Ala Cys
20 25 30
Asn Asp Met Thr Pro Glu Gln Met Ala Thr Asn Val Asn Cys Ser Ser
35 40 45
Pro Glu Arg His Thr Arg Ser Tyr Asp Tyr Met Glu Gly Gly Asp Ile
50 55 60

```
Arg Val Arg Arg Leu Phe Cys Arg Thr Gln Trp Tyr Leu Arg Ile Asp
65                  70                  75                  80
Lys Arg Gly Lys Val Lys Gly Thr Gln Glu Met Lys Asn Asn Tyr Asn
                85                  90                  95
Ile Met Glu Ile Arg Thr Val Ala Val Gly Ile Val Ala Ile Lys Gly
            100                 105                 110
Val Glu Ser Glu Phe Tyr Leu Ala Met Asn Lys Glu Gly Lys Leu Tyr
        115                 120                 125
Ala Lys Lys Glu Cys Asn Glu Asp Cys Asn Phe Lys Glu Leu Ile Leu
    130                 135                 140
Glu Asn His Tyr Asn Thr Tyr Ala Ser Ala Lys Trp Thr His Asn Gly
145                 150                 155                 160
Gly Glu Met Phe Val Ala Leu Asn Gln Lys Gly Ile Pro Val Arg Gly
                165                 170                 175
Lys Lys Thr Lys Lys Glu Gln Lys Thr Ala His Phe Leu Pro Met Ala
            180                 185                 190
Ile Thr
```

<210> 6
<211> 207
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Ala Glu Val Gly Gly Val Phe Ala Ser Leu Asp Trp Asp Leu His
 1               5                  10                  15
Gly Phe Ser Ser Ser Leu Gly Asn Val Pro Leu Ala Asp Ser Pro Gly
            20                  25                  30
Phe Leu Asn Glu Arg Leu Gly Gln Ile Glu Gly Lys Leu Gln Arg Gly
        35                  40                  45
Ser Pro Thr Asp Phe Ala His Leu Lys Gly Ile Leu Arg Arg Arg Gln
    50                  55                  60
Leu Tyr Cys Arg Thr Gly Phe His Leu Glu Ile Phe Pro Asn Gly Thr
65                  70                  75                  80
Val His Gly Thr Arg His Asp His Ser Arg Phe Gly Ile Leu Glu Phe
                85                  90                  95
Ile Ser Leu Ala Val Gly Leu Ile Ser Ile Arg Gly Val Asp Ser Gly
            100                 105                 110
Leu Tyr Leu Gly Met Asn Glu Arg Gly Glu Leu Tyr Gly Ser Lys Lys
        115                 120                 125
Leu Thr Arg Glu Cys Val Phe Arg Glu Gln Phe Glu Glu Asn Trp Tyr
    130                 135                 140
```

```
Asn Thr Tyr Ala Ser Thr Leu Tyr Lys His Ser Asp Ser Glu Arg Gln
145                 150                 155                 160
Tyr Tyr Val Ala Leu Asn Lys Asp Gly Ser Pro Arg Glu Gly Tyr Arg
                165                 170                 175
Thr Lys Arg His Gln Lys Phe Thr His Phe Leu Pro Arg Pro Val Asp
            180                 185                 190
Pro Ser Lys Leu Pro Ser Met Ser Arg Asp Leu Phe His Tyr Arg
        195                 200                 205
```

<210> 7
<211> 208
<212> PRT
<213> Homo sapiens

<400> 7

```
Met Ala Pro Leu Gly Glu Val Gly Asn Tyr Phe Gly Val Gln Asp Ala
 1               5                  10                  15
Val Pro Phe Gly Asn Val Pro Val Leu Pro Val Asp Ser Pro Val Leu
            20                  25                  30
Leu Ser Asp His Leu Gly Gln Ser Glu Ala Gly Gly Leu Pro Arg Gly
        35                  40                  45
Pro Ala Val Thr Asp Leu Asp His Leu Lys Gly Ile Leu Arg Arg Arg
    50                  55                  60
Gln Leu Tyr Cys Arg Thr Gly Phe His Leu Glu Ile Phe Pro Asn Gly
65                  70                  75                  80
Thr Ile Gln Gly Thr Arg Lys Asp His Ser Arg Phe Gly Ile Leu Glu
                85                  90                  95
Phe Ile Ser Ile Ala Val Gly Leu Val Ser Ile Arg Gly Val Asp Ser
            100                 105                 110
Gly Leu Tyr Leu Gly Met Asn Glu Lys Gly Glu Leu Tyr Gly Ser Glu
        115                 120                 125
Lys Leu Thr Gln Glu Cys Val Phe Arg Glu Gln Phe Glu Glu Asn Trp
    130                 135                 140
Tyr Asn Thr Tyr Ser Ser Asn Leu Tyr Lys His Val Asp Thr Gly Arg
145                 150                 155                 160
Arg Tyr Tyr Val Ala Leu Asn Lys Asp Gly Thr Pro Arg Glu Gly Thr
                165                 170                 175
Arg Thr Lys Arg His Gln Lys Phe Thr His Phe Leu Pro Arg Pro Val
            180                 185                 190
Asp Pro Asp Lys Val Pro Glu Leu Tyr Lys Asp Ile Leu Ser Gln Ser
        195                 200                 205
```

<210> 8

<211> 239
<212> PRT
<213> Homo sapiens

<400> 8

```
Met Gly Leu Ile Trp Leu Leu Leu Leu Ser Leu Leu Glu Pro Gly Trp
 1               5                  10                  15
Pro Ala Ala Gly Pro Gly Ala Arg Leu Arg Arg Asp Ala Gly Gly Arg
            20                  25                  30
Gly Gly Val Tyr Glu His Leu Gly Gly Ala Pro Arg Arg Arg Lys Leu
        35                  40                  45
Tyr Cys Ala Thr Lys Tyr His Leu Gln Leu His Pro Ser Gly Arg Val
    50                  55                  60
Asn Gly Ser Leu Glu Asn Ser Ala Tyr Ser Ile Leu Glu Ile Thr Ala
65                  70                  75                  80
Val Glu Val Gly Ile Val Ala Ile Arg Gly Leu Phe Ser Gly Arg Tyr
                85                  90                  95
Leu Ala Met Asn Lys Arg Gly Arg Leu Tyr Ala Ser Glu His Tyr Ser
            100                 105                 110
Ala Glu Cys Glu Phe Val Glu Arg Ile His Glu Leu Gly Tyr Asn Thr
        115                 120                 125
Tyr Ala Ser Arg Leu Tyr Arg Thr Val Ser Ser Thr Pro Gly Ala Arg
    130                 135                 140
Arg Gln Pro Ser Ala Glu Arg Leu Trp Tyr Val Ser Val Asn Gly Lys
145                 150                 155                 160
Gly Arg Pro Arg Arg Gly Phe Lys Thr Arg Arg Thr Gln Lys Ser Ser
                165                 170                 175
Leu Phe Leu Pro Arg Val Leu Asp His Arg Asp His Glu Met Val Arg
            180                 185                 190
Gln Leu Gln Ser Gly Leu Pro Arg Pro Pro Gly Lys Gly Val Gln Pro
        195                 200                 205
Arg Arg Arg Arg Gln Lys Gln Ser Pro Asp Asn Leu Glu Pro Ser His
    210                 215                 220
Val Gln Ala Ser Arg Leu Gly Ser Gln Leu Glu Ala Ser Ala His
225                 230                 235
```

<210> 9
<211> 206
<212> PRT
<213> Homo sapiens

<400> 9

```
Met Ser Gly Pro Gly Thr Ala Ala Val Ala Leu Leu Pro Ala Val Leu
 1               5                  10                  15
Leu Ala Leu Leu Ala Pro Trp Ala Gly Arg Gly Gly Ala Ala Ala Pro
            20                  25                  30
Thr Ala Pro Asn Gly Thr Leu Glu Ala Glu Leu Glu Arg Arg Trp Glu
        35                  40                  45
Ser Leu Val Ala Leu Ser Leu Ala Arg Leu Pro Val Ala Ala Gln Pro
    50                  55                  60
Lys Glu Ala Ala Val Gln Ser Gly Ala Gly Asp Tyr Leu Leu Gly Ile
65                  70                  75                  80
Lys Arg Leu Arg Arg Leu Tyr Cys Asn Val Gly Ile Gly Phe His Leu
                85                  90                  95
Gln Ala Leu Pro Asp Gly Arg Ile Gly Gly Ala His Ala Asp Thr Arg
            100                 105                 110
Asp Ser Leu Leu Glu Leu Ser Pro Val Glu Arg Gly Val Val Ser Ile
        115                 120                 125
Phe Gly Val Ala Ser Arg Phe Phe Val Ala Met Ser Ser Lys Gly Lys
    130                 135                 140
Leu Tyr Gly Ser Pro Phe Phe Thr Asp Glu Cys Thr Phe Lys Glu Ile
145                 150                 155                 160
Leu Leu Pro Asn Asn Tyr Asn Ala Tyr Glu Ser Tyr Lys Tyr Pro Gly
                165                 170                 175
Met Phe Ile Ala Leu Ser Lys Asn Gly Lys Thr Lys Lys Gly Asn Arg
            180                 185                 190
Val Ser Pro Thr Met Lys Val Thr His Phe Leu Pro Arg Leu
        195                 200                 205
```

<210> 10
<211> 225
<212> PRT
<213> Homo sapiens

<400> 10

```
Met Ala Ala Leu Ala Ser Ser Leu Ile Arg Gln Lys Arg Glu Val Arg
 1               5                  10                  15
Glu Pro Gly Gly Ser Arg Pro Val Ser Ala Gln Arg Arg Val Cys Pro
            20                  25                  30
Arg Gly Thr Lys Ser Leu Cys Gln Lys Gln Leu Leu Ile Leu Leu Ser
        35                  40                  45
Lys Val Arg Leu Cys Gly Gly Arg Pro Ala Arg Pro Asp Arg Gly Pro
    50                  55                  60
Glu Pro Gln Leu Lys Gly Ile Val Thr Lys Leu Phe Cys Arg Gln Gly
65                  70                  75                  80
```

Phe Tyr Leu Gln Ala Asn Pro Asp Gly Ser Ile Gln Gly Thr Pro Glu
85 90 95
Asp Thr Ser Ser Phe Thr His Phe Asn Leu Ile Pro Val Gly Leu Arg
100 105 110
Val Val Thr Ile Gln Ser Ala Lys Leu Gly His Tyr Met Ala Met Asn
115 120 125
Ala Glu Gly Leu Leu Tyr Ser Ser Pro His Phe Thr Ala Glu Cys Arg
130 135 140
Phe Lys Glu Cys Val Phe Glu Asn Tyr Tyr Val Leu Tyr Ala Ser Ala
145 150 155 160
Leu Tyr Arg Gln Arg Arg Ser Gly Arg Ala Trp Tyr Leu Gly Leu Asp
165 170 175
Lys Glu Gly Gln Val Met Lys Gly Asn Arg Val Lys Lys Thr Lys Ala
180 185 190
Ala Ala His Phe Leu Pro Lys Leu Leu Glu Val Ala Met Tyr Gln Glu
195 200 205
Pro Ser Leu His Ser Val Pro Glu Ala Ser Pro Ser Ser Pro Pro Ala
210 215 220
Pro
225

<210> 11
<211> 268
<212> PRT
<213> Homo sapiens

<400> 11
Met Ser Leu Ser Phe Leu Leu Leu Leu Phe Phe Ser His Leu Ile Leu
1 5 10 15
Ser Ala Trp Ala His Gly Glu Lys Arg Leu Ala Pro Lys Gly Gln Pro
20 25 30
Gly Pro Ala Ala Thr Asp Arg Asn Pro Ile Gly Ser Ser Ser Arg Gln
35 40 45
Ser Ser Ser Ser Ala Met Ser Ser Ser Ser Ala Ser Ser Ser Pro Ala
50 55 60
Ala Ser Leu Gly Ser Gln Gly Ser Gly Leu Glu Gln Ser Ser Phe Gln
65 70 75 80
Trp Ser Pro Ser Gly Arg Arg Thr Gly Ser Leu Tyr Cys Arg Val Gly
85 90 95
Ile Gly Phe His Leu Gln Ile Tyr Pro Asp Gly Lys Val Asn Gly Ser
100 105 110
His Glu Ala Asn Met Leu Ser Val Leu Glu Ile Phe Ala Val Ser Gln
115 120 125

```
Gly Ile Val Gly Ile Arg Gly Val Phe Ser Asn Lys Phe Leu Ala Met
    130                 135                 140
Ser Lys Lys Gly Lys Leu His Ala Ser Ala Lys Phe Thr Asp Asp Cys
145                 150                 155                 160
Lys Phe Arg Glu Arg Phe Gln Glu Asn Ser Tyr Asn Thr Tyr Ala Ser
                165                 170                 175
Ala Ile His Arg Thr Glu Lys Thr Gly Arg Glu Trp Tyr Val Ala Leu
            180                 185                 190
Asn Lys Arg Gly Lys Ala Lys Arg Gly Cys Ser Pro Arg Val Lys Pro
        195                 200                 205
Gln His Ile Ser Thr His Phe Leu Pro Arg Phe Lys Gln Ser Glu Gln
    210                 215                 220
Pro Glu Leu Ser Phe Thr Val Thr Val Pro Glu Lys Lys Asn Pro Pro
225                 230                 235                 240
Ser Pro Ile Lys Ser Lys Ile Pro Leu Ser Ala Pro Arg Lys Asn Thr
                245                 250                 255
Asn Ser Val Lys Tyr Arg Leu Lys Phe Arg Phe Gly
            260                 265
```

<210> 12
<211> 247
<212> PRT
<213> Homo sapiens

<400> 12

```
Met Ala Ala Ala Ile Ala Ser Gly Leu Ile Arg Gln Lys Arg Gln Ala
 1               5                  10                  15
Arg Glu Gln His Trp Asp Arg Pro Ser Ala Ser Arg Arg Arg Ser Ser
            20                  25                  30
Pro Ser Lys Asn Arg Gly Leu Cys Asn Gly Asn Leu Val Asp Ile Phe
        35                  40                  45
Ser Lys Val Arg Ile Phe Gly Leu Lys Lys Arg Arg Leu Arg Arg Gln
    50                  55                  60
Asp Pro Gln Leu Lys Gly Ile Val Thr Arg Leu Tyr Cys Arg Gln Gly
65                  70                  75                  80
Tyr Tyr Leu Gln Met His Pro Asp Gly Ala Leu Asp Gly Thr Lys Asp
                85                  90                  95
Asp Ser Thr Asn Ser Thr Leu Phe Asn Leu Ile Pro Val Gly Leu Arg
            100                 105                 110
Val Val Ala Ile Gln Gly Val Lys Thr Gly Leu Tyr Ile Ala Met Asn
        115                 120                 125
Gly Glu Gly Tyr Leu Tyr Pro Ser Glu Leu Phe Thr Pro Glu Cys Lys
    130                 135                 140
```

```
Phe Lys Glu Ser Val Phe Glu Asn Tyr Tyr Val Ile Tyr Ser Ser Met
145                 150                 155                 160
Leu Tyr Arg Gln Gln Glu Ser Gly Arg Ala Trp Phe Leu Gly Leu Asn
                165                 170                 175
Lys Glu Gly Gln Ala Met Lys Gly Asn Arg Val Lys Lys Thr Lys Pro
            180                 185                 190
Ala Ala His Phe Leu Pro Lys Pro Leu Glu Val Ala Met Tyr Arg Glu
        195                 200                 205
Pro Ser Leu His Asp Val Gly Glu Thr Val Pro Lys Pro Gly Val Thr
    210                 215                 220
Pro Ser Lys Ser Thr Ser Ala Ser Ala Ile Met Asn Gly Gly Lys Pro
225                 230                 235                 240
Val Asn Lys Ser Lys Thr Thr
                245
```

<210> 13
<211> 168
<212> PRT
<213> Homo sapiens

<400> 13

```
Met Ala Ser Lys Glu Pro Gln Leu Lys Gly Ile Val Thr Arg Leu Phe
 1               5                  10                  15
Ser Gln Gln Gly Tyr Phe Leu Gln Met His Pro Asp Gly Thr Ile Asp
            20                  25                  30
Gly Thr Lys Asp Glu Asn Ser Asp Tyr Thr Leu Phe Asn Leu Ile Pro
        35                  40                  45
Val Gly Leu Arg Val Val Ala Ile Gln Gly Val Lys Ala Ser Leu Tyr
    50                  55                  60
Val Ala Met Asn Gly Glu Gly Tyr Leu Tyr Ser Ser Asp Val Phe Thr
65                  70                  75                  80
Pro Glu Cys Lys Phe Lys Glu Ser Val Phe Glu Asn Tyr Tyr Val Ile
                85                  90                  95
Tyr Ser Ser Thr Leu Tyr Arg Gln Gln Glu Ser Gly Arg Ala Trp Phe
            100                 105                 110
Leu Gly Leu Asn Lys Glu Gly Gln Ile Met Lys Gly Asn Arg Val Glu
        115                 120                 125
Lys Thr Lys Pro Ser Ser His Phe Val Pro Lys Pro Ile Glu Val Cys
    130                 135                 140
Met Tyr Arg Glu Pro Ser Leu His Glu Ile Gly Glu Asn Lys Gly Val
145                 150                 155                 160
Gln Gly Lys Phe Trp Thr Pro Pro
                165
```

<210> 14
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Motif

<221> VARIANT
<222> (2)...(2)
<223> Xaa is any amino acid residue except cysteine.

<221> VARIANT
<222> (4)...(4)
<223> Xaa is any amino acid residue except cysteine.

<400> 14
Cys Xaa Phe Xaa Glu
1 5

<210> 15
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> peptide

<400> 15
Glu Tyr Pro Met Glu
1 5